VOL. 14 No. 2, pp. 33-68 AUGUST 1952

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

ELECTRONIC INTENSIFICATION OF FLUOROSCOPIC IMAGES

by M. C. TEVES and T. TOL.

621.386.8:616-073.75:621.383.8

In essence, the generation of X-rays, which entails the production of a beam of electrons in a vacuum tube, falls within the technique of electronics. It is a noteworthy point that in X-ray diagnostics the subsequent stage, that is, the conversion to a visible picture of the X-rays which have passed through the object to be examined, is now also tending to become dependent on electronics for a satisfactory solution. In this context it is found that the number of X-ray quanta available virtually sets a limit to the amount of detail in the information that can be derived from fluoroscopy.

The problem of screen brightness in fluoroscopy

X-rays have the capacity to damage or destroy human tissue. In X-ray therapy, use is made of this fact by exposing morbid tissue to these rays for such periods and at such intensities that in total it will receive a certain, heavy, dose. In X-ray diagnostics, on the other hand, care must be taken that patients receive the smallest possible dose during examination so that they will not suffer any injurious consequences, even as a result of repeated examinations.

Using the conventional method of fluoroscopy, therefore, the radiologist must do the best he can with an X-ray image of very low brightness. In fact, for a given duration of an examination (a few minutes at the most), limitation of the total dosage means limitation of the dosage rate to which the patient may be subjected and, accordingly — dependent upon the absorption of the rays by the patient — limitation of the brightness that can be obtained from the screen. In gastric fluoroscopy, for example, with the X-ray tube operating on a voltage of, say, 90 kV and a current of 2 mA, the image obtained on a high-quality screen has an average brightness of about 0.003 cd/m^2. For comparative purposes it may be said that the average brightness of a landscape by the light of the full moon is 0.01 cd/m^2.

In order to be able to make a good visual assessment at such low brightness values, the radiologist must first remain in darkness for at least 15 minutes, so that his eyes may be adapted to the low brightness level. Even so, his visual acuity and contrast sensitivity ¹) are then fairly low. Hence, for the diagnosis of conditions involving slight abnormalities which often reveal only a low contrast, this is a very inconvenient limitation. It will therefore be easily appreciated that, since the inception of X-ray diagnostics, higher image brightnesses have been an ever-present need.

If more information is to be obtained from the fluoroscopic screen and at the same time a solution found for the problem of the necessity for dark adaptation of the eyes, an increase in image brightness to the extent of at least a factor of 100 to 1000 is needed, as pointed out by **Chamberlain** in 1942 ²). It would not be possible to achieve this simply by increasing the efficacy of existing types of screen; these convert roughly 5% of the X-rays falling on them to visible radiation, and this radiation has approximately the most advantageous wavelength (i.e. in the yellow-green part of the spectrum, where, for a given strength of the radia-

---

¹) Visual acuity is defined as $1/a$, where a is the angle subtended by the finest distinguishable detail. Contrast sensitivity is defined as $1/c$, where c is the lowest distinguishable contrast $\Delta B/B$; in this, B is the brightness of the background and $B \pm \Delta B$ the brightness of an object seen against that background.

²) W. E. Chaimberlain, Fluoroscopes and fluoroscopy, Radiology **38**, 383-413, 1942.

tion, the strongest impression of brightness is received by the eyes). Thus, the gain that is theoretically possible in this way would not exceed a factor of 20.

In recent years several methods, using electronic aids, have in fact been devised for intensifying the brightness of the fluoroscopic image to a considerable degree [3]. In the present article we shall describe an image intensifier designed in the Philips Laboratories at Eindhoven. This equipment is based on the well-known principle of the image converter, another development of this laboratory, a description of which was given as early as in 1934 [4]. The fluoroscopic images obtained with the aid of this instrument are roughly 1000 times brighter than those produced on the ordinary fluoroscopic screen.

Description of the intensifier

The more important details of the unit are illustrated in *fig. 1.* A fluoroscopic screen on a support of thin aluminium is mounted in an evacuated glass tube. A photocathode lies in contact with the screen. When X-rays pass through the glass wall of the tube to the screen, this fluoresces. The fluorescent light sets electrons free from the photocathode; at any point on the photocathode the number of photo-electrons released per second is proportional to the luminous intensity of the fluoroscopic screen at that point. The luminous image with all its variations in brightness is thus transformed into an "electronic image" with corresponding variations in current density. By means of an electric field of a certain configuration (electrostatic "electronic lens") this electronic image is reproduced on a second fluorescent screen, the viewing screen, reduced 9 times in size. The energy of the electrons falling on this screen is in part converted to fluorescent light, and a reproduction of the image on the fluoroscopic screen is thus seen on the viewing screen reduced 9 times. This reduced image is viewed through a simple microscope of about 9 × magnification, and finally the fluoroscopic image is thus seen erect and in its original size (both the electronic lens and the microscope produce inverted images).

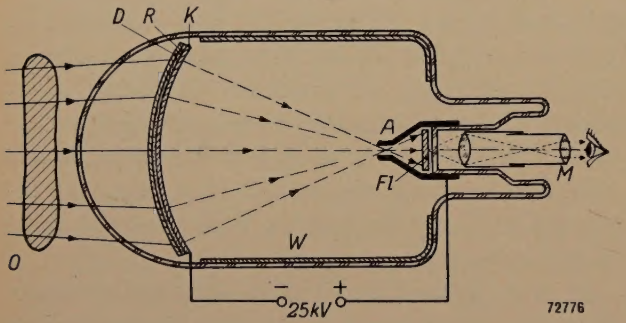


Fig. 1. Diagrammatic cross-section of the image intensifier. *R* = fluoroscopic screen on which the X-rays are received after passing through the object to be examined *O*, and the glass wall of the tube. *D* = support of fluoroscopic screen and photocathode *K*. The fluorescent radiations produced in *R* release electrons from the photocathode *K*. The "electronic image" is reproduced on a reduced scale by the electric field between *K* and the perforated anode *A*, on a viewing screen *Fl*; this image is observed through a simple microscope *M*. *W* is a conductive layer on the inside of the tube.

The image seen through the microscope is thus identical with that seen on an ordinary fluoroscopic screen; it is, however, about 1000 times brighter, assuming that the image intensifier and the fluoroscopic screen both receive X-rays of the same intensity.

How is the brightness intensification obtained?

Two factors (which are not in fact independent of each other) contribute towards the intensification of the image brightness.

The first and more readily understood factor is an increase in the total luminous flux. The electrons emitted by the photocathode are not only focused on the viewing screen by the electrical field, they are also accelerated by it; in the particular instrument under review a potential of 25 kV is applied between the viewing screen and the photocathode. The more energy possessed by an electron on arrival at the screen, the more fluorescent light it will produce. Notwithstanding the fact that only about 1 in 10 of the light quanta from the fluoroscopic screen liberates a photo-electron and that only some 10% of the electronic energy is converted to light by the viewing screen, the last mentioned screen — owing to the energy imparted to the electrons — is able to yield a total of 10 to 15 times as much luminous flux (in lumens) as an ordinary fluoroscopic screen, for the same object viewed.

The second, more important factor is the gain produced by the above-mentioned electron-optical reduction of the size of the image. All the electrons coming from the photocathode contribute towards the final image, and the energy that each electron is capable of converting to light after passing through the electron-optical system is independent of whether the electrons are distributed over a large or over a small area. By employing a reduction of 9 times we concentrate the electrons in an area 9^2 times smaller, and the total luminous flux is thus emitted from an area that is about 80 times smaller than would be the case in reproduction on a scale of 1 : 1. This, by definition, represents an increase in brightness [5] by a factor of 80.

---

[3] J. W. Coltman, Fluoroscopic image brightening by electronic means, Radiology **51**, 359-367, 1948.
R. J. Moon, Amplifying and intensifying fluoroscopic images by means of a scanning X-ray tube, Science **112**, 389-395, 1950.
R. H. Morgan and R. E. Sturm, The John Hopkins fluoroscopic screen intensifier, Radiology **57**, 556-560, 1951.
P. C. Hodges and L. Staggs, Electronic amplification of the röntgenoscopic image, Amer. J. Röntgenology Rad. Ther., **66**, 705-710, 1951.

[4] G. Holst, J. H. de Boer, M. C. Teves and C. F. Veenemans, An apparatus for the transformation of light of long wavelength into light of short wavelength, Physica **1**, 297-305, 1934.

The total gain in brightness is equal to the product of the "lumen gain" and the gain due to the reduction in size, i.e. in the tube in question, 10 to 15 times $9^2 = 800$ to 1200.

It is perhaps not superfluous to add that the gain brought about by the electron-optical reduction cannot be obtained by the ordinary methods of optical reduction. According to Abbe's law, the apparent brightness of an object does not vary as a result of the introduction of any kind of optical instrument between the object and the eye of the observer, that is, if such an instrument can be regarded as ideal, without the absorption and reflection losses that normally occur in lenses and mirrors (and assuming the exit pupil of the instrument to be at least as large as the pupil of the eye).

This will readily be appreciated on reference to *fig. 2a*. Let the brightness of the object be B, and the solid angle subtended by the lens when seen from the object ω. The lens then receives from a surface element df a luminous flux of $\Phi = B\omega df$. This luminous flux is concentrated in a smaller area df' in the image space. This area thus again emits the luminous flux Φ, this time, however, within the greater solid angle ω' (or, with diffusion, in an even larger angle). The brightness of df' is $B' = \Phi/\omega' df'$ and, since we know from the laws of optics that $\omega' df' = \omega df$ (the sine law), B' is equal to B.

The difference that is seen to exist in this case by no means implies that the analogy between electron-optics and light-optics is no longer valid. Abbe's law as stated in the above simplified form holds good only when the object and the image are located in media having the same refractive index, and in ordinary optical instruments this is usually the case. If the refractive index of the medium containing the image is higher than that of the object, the brightness of the image can certainly be greater than that of the object itself. This is exactly what happens in electron-optics. The potential field in which the electrons are moving may be regarded as a medium the refractive index of which is continuously variable; in fact, the index is at any point proportional to the velocity of the electrons. The refractive index at the site of the viewing screen is therefore about 100 times higher than that at the photocathode.

The analogy is thus not only restored formally, but may again be illustrated in a simple way. Owing to the continuous variability of the refractive index, the "rays" are now not straight, but curved. The path of the rays in the electrostatic electron-lens therefore takes the form shown in fig. 2*b*, in contrast with fig. 2*a*; it will be seen that the flux radiated from the surface element df within a solid angle almost equal to 2π is employed and is concentrated in a much smaller solid angle at the surface df'.

By reason of the influence of the velocity of the electrons, it is seen that the "lumen intensification" and the brightness gain due to the reduction in size virtually spring from the same cause, i.e. the presence of the accelerating field.

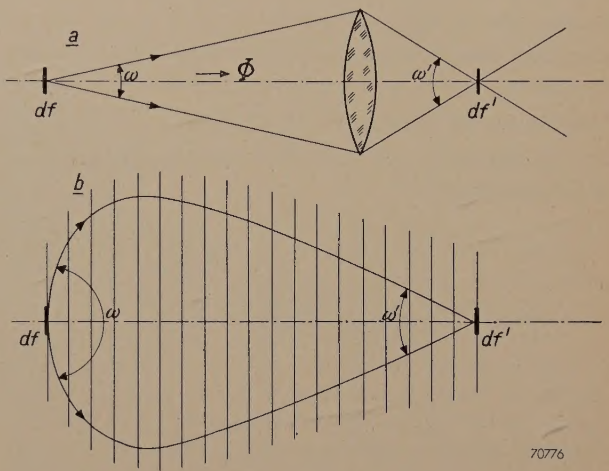


Fig. 2. *a*) Path of rays on optical reproduction of a surface df, at df' (illustrating Abbe's law).
b) The same, as applicable to electron-optical reproduction with an electrostatic lens, as in the image intensifier.

In the light of Abbe's law it will now be understood why in the last stage of the image intensifier, viz. the optical enlargement by the microscope of the image on the viewing screen, the gain in brightness arising from the electron-optical reduction is not neutralised (provided that the microscope meets the condition as regards the exit pupil to which reference has already been made).

It may possibly strike the reader that in the foregoing, particularly in our explanation of the lumen gain, the luminous flux of the viewing screen is each time compared with that of an ordinary fluoroscopic screen and not, as might appear more obvious, with the luminous flux of the fluoroscopic screen in the image intensifier itself. The reason is that the last-mentioned screen need not be observed and that the luminosity curve of the fluorescent light emitted by it may therefore be such as to be quite unsuitable from the point of view of direct observation; in fact the fluorescent radiation might even be entirely invisible, for instance ultra-violet or infra-red, provided only that the photocathode be sensitive to such radiation. Our image intensifier is equipped with a photocathode the quantum efficiency of which is highest (about 10%) when it is exposed to blue light. A phosphor giving blue fluorescence was accordingly chosen for the fluoroscopic screen. For the viewing screen, on the contrary, if a choice from different materials of the same luminous efficiency is available, that colour should be employed of which the fluorescent spectrum coincides as nearly as possible with the maximum in the spectral sensitivity curve of the eye. This would be a phosphor emitting yellow-green light. Clearly, if we compared the brightness of such a viewing screen with that of the blue emission of the fluoroscopic screen in the image intensifier, we would calculate a much greater brightness intensification value than 1000 times, but at the same time one that would have no practical significance.

---

[5] According to a recent standardisation decision on lighting technology terms, we should speak here of luminance.

For this reason it would be as well to dwell for a moment on the precise meaning of the name "image intensifier" or "brightness intensifier". In effect the instrument is an image converter, which converts the "image" of the invisible X-rays to a visible image, by analogy with the well-known instruments for transforming infra-red radiations into visible images (see the article referred to in note [4]) which are employed in numerous different forms. The fact that this conversion in our case involves as intermediate stage a visible fluorescent image (as in normal fluoroscopy) is attributable only to the impossibility of liberating photo-electrons direct from a photocathode with any degree of efficiency (and with not too high an initial velocity) by means of the X-rays themselves. Reasonable efficiency in this release of electrons is naturally of the highest importance in the fluoroscopic image intensifier, as otherwise the effect of the subsequent gain would be lost and we might then just as well employ only the "image conversion" as obtained from the ordinary fluoroscopic screen.

Details of construction

The blue fluorescence of the fluoroscopic screen as matched to the photocathode is obtained from zinc sulphide activated with silver, and the yellow-green fluorescing viewing screen is prepared from very finely divided zinc sulphide-zinc selenide. The effective diameter of the cathode face is 13.5 cm, this thus representing the maximum size of the object to be examined. The image on the viewing screen is only 1/9th of this size, viz. 1.5 cm across.

The electrical field producing the electron-optical image is generated basically by means of only two electrodes, one of which (the cathode) is formed by the concave spherical surface of the combined fluoroscopic screen and photocathode, as well as by a conductive layer on the inside wall of the glass tube. The other electrode (the anode) is also spherically curved and is concentric with the photocathode; it has in it an aperture to allow the electrons to pass to the viewing screen behind it [6]). Between the cathode and the anode a potential of 25 kV is applied.

This simple design, employing only two electrodes, has proved to be very effective. Definition in the image is ample and the field strength on the only slightly curved photocathode is very low, this being an asset for a reason which is given in another paragraph. When the tube was still in the experimental stage, the configuration offered sufficient latitude, viz, in the radii of curvature of cathode and anode, allowing for example a variation in the amount of reduction of the image. For practical purposes it was later found to be better for the inner lining of the tube to function as a separate electrode and to maintain it at a slightly positive potential with respect to the photocathode. Without this, only the electrons emitted from the central part of the photocathode pass through the opening in the anode, the peripheral parts thus not being reproduced. With the auxiliary potential applied the peripheral electrons also pass through the anode, the effective image area of the fluoroscopic screen being thus almost doubled. At the same time, the slight "pin-cushion" distortion of the image that occurs without the auxiliary voltage is in this way almost entirely eliminated. Another important advantage of the auxiliary potential — which is variable — is that by means of it the image on the viewing screen can be adjusted for maximum definition whilst the tube is in use. Consequently, the tolerances in the electrode spacing, during the manufacture of the glass tube, can be relatively wide.

The 25 kV applied between cathode and anode may be obtained for instance from the well-known small high-tension generator developed in this Laboratory for projection television [7]). This generator also supplies the necessary voltage for the inner tube coating. The cathode is earthed, so that only one of the three poles of the image intensifier need be suitable for carrying high voltage. Only a very small amount of power is required to operate the tube, the photo-current for normal fluoroscopy being in the order of 10^{-9} A. For this reason and also in view of the fact that in this case no stringent conditions need be imposed as regards voltage stability (a basic advantage in purely electrostatic electron-optical systems), an even simpler high-tension unit would suffice.

---

[6]) For a more detailed description of the principle of this two-electrode lens the reader is referred to an article by P. Schagen, H. Bruining and J. C. Francken, Philips Research Reports 7, 119-130, 1952 (No. 2).

[7]) G. J. Siezen and F. Kerkhof, Philips Techn. Rev. **10**, 125-134, 1948.

Fig. 3 depicts the complete image amplifier tube. It is cylindrical, the diameter being 17.5 cm; the over-all length, including the small microscope, is 45 cm. To protect the tube against jarring and to afford local protection from stray X-rays, the tube proper is contained in a lead-lined aluminium jacket which can be mounted quite simply on a stand; see *fig. 4*. Tube and jacket together weigh 7.5 kg approx.

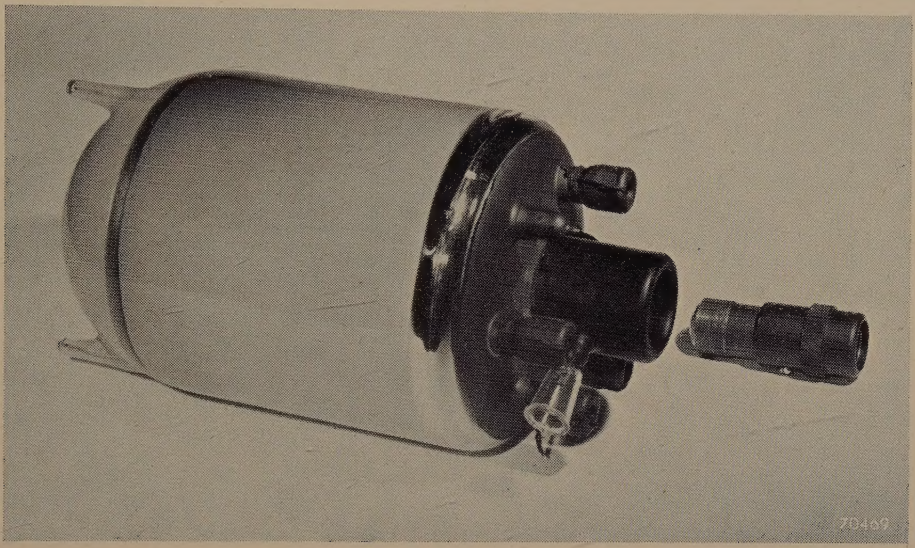


Fig. 3. The Philips X-ray image intensifier, producing an image that is approximately 1000 times brighter than that of the ordinary fluoroscopic screen. The tube is 45 cm in length, 17.5 cm in diameter. The effective area of the screen is a circle 13.5 cm in diameter. On the right will be seen the viewing end into which the microscope shown alongside is fitted.

Definition and contrast in the image

When the image intensifier is in use it is not necessary for the radiologist to allow time for his eyes to adapt themselves; examinations can be made in a room with almost normal illumination. This is in itself a great practical advantage that will almost certainly outweigh the drawback that the image cannot be surveyed with the naked eye, but only through an optical system. Moreover, it may be expected, as explained in the introduction, that the radiologist will be in a position to perceive smaller details than is possible with the ordinary fluoroscopic screen because of the higher visual acuity and contrast sensitivity of the eye at the increased brightness level. In order that this expectation — which we shall enlarge upon presently — may be realised, it is an essential condition that the blurring and loss of contrast introduced by the image intensifier itself, as by any other form of optical instrument, shall be kept within certain limits.

Definition

In normal fluoroscopy the definition of the image will be determined by the finite dimensions of the focus of the X-ray tube (half-shadow width or geometrical blurring) as well as by the thickness and grain size of the fluoroscopic screen (intrinsic or screen blurring). Screen blurring is usually about 0.7 mm, sometimes less, down to 0.4 mm; the geometrical blurring is generally slightly less. The resulting definition in the fluoroscopic image is thus somewhat less than 1 mm (the total blurring being equal, approximately, to the square root of the sum of the squares of the geometric blurring and the screen blurring [8])).

Screen blurring can be reduced considerably by using thinner screens (and finer grain). This is not done in the case of ordinary fluoroscopy, as such screens would transmit a larger amount of X-rays, which are thus ineffective, resulting in reduced screen efficiency [9]). Moreover, better definition would be of little use, since, at the low brightness levels of normal fluoroscopic screens, visual acuity is in any case too low for the perception of finer detail.

---

[8]) See H. A. Klasens, The blurring of X-ray images, Philips Techn. Rev. 9, 364-369, 1948.

[9]) As far as efficiency is concerned there is an optimum thickness, for in screens that are too thick, apparently too much absorption of the emitted light will take place in the screen itself. (This conflict between useful absorption of the incident radiation and wasteful absorption of the rays emitted in the image intensifier again occurs in the photocathode and in the viewing screen.)

In the image intensifier the position is very different. Here, too, there is a primary lack of definition in the image owing to the intrinsic blurring in the fluoroscopic screen employed. However, there is now no objection to the use of a relatively thin screen, thus sacrificing some of the gain in brightness to benefit the definition. A real advantage is obtained by so doing, for the visual acuity at the brightness level of the image intensifier is so high that the eye can reap the benefit of the increased definition.

For this reason a thinner fluoroscopic screen is used in the image intensifier than is provided for ordinary fluoroscopy. On the other hand the reduction in the screen thickness cannot be carried too far, for a reason other than that connected with efficiency, as will be seen presently.

In the case of the image intensifier there are also other possible causes of blurring. These must of course not detract from the available gain in definition, and in fact, in our model, they do not. The grain size of the viewing screen used is very fine and the screen itself very thin (the penetration depth of electrons at 25 kV is about 12 μ), so that the intrinsic blurring in this screen, referred to natural image size, is less than that of the fluoroscopic screen. Blurring by the electron-optical reproduction system is at least one order of size less (see formula (26) in the article referred to in note [6])) and can therefore be ignored, as can, also, the limitation in the resolving power of the microscope through which the screen is viewed. No appreciable blurring occurs in the photocathode either, since this is in direct contact with the fluoroscopic screen, and, in order to avoid too much absorption of the liberated photo-electrons, is extremely thin.

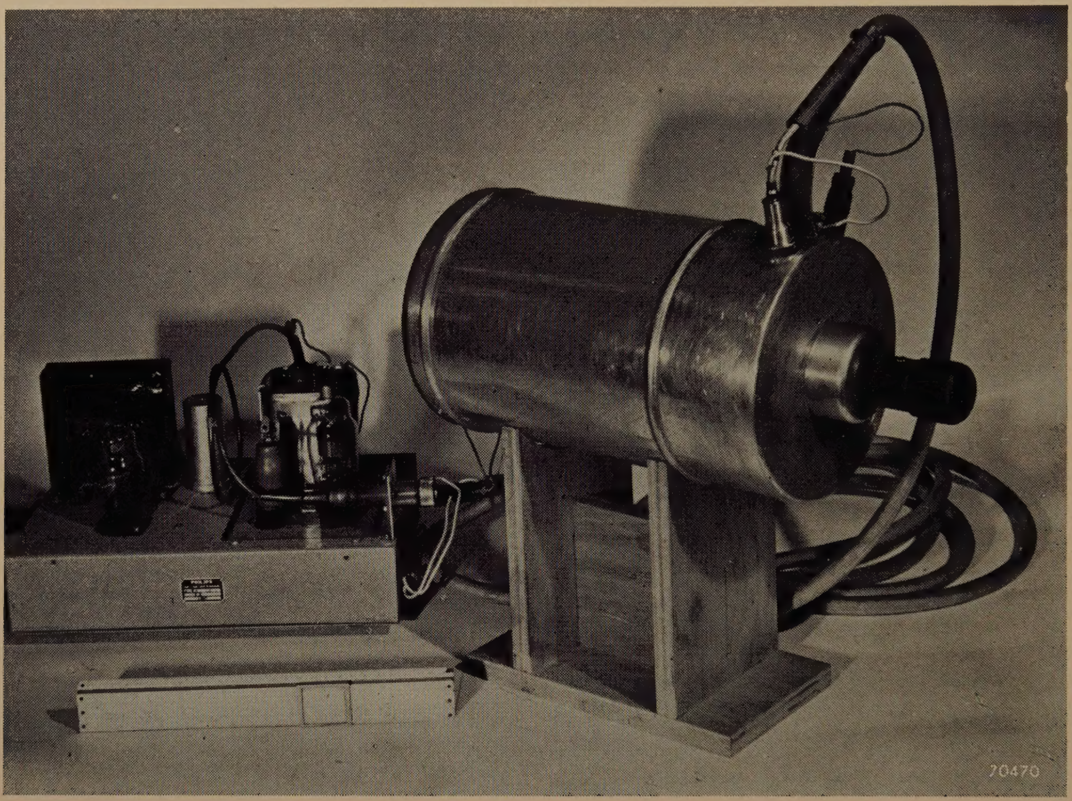


Fig. 4. The complete equipment for fluoroscopy by means of the image intensifier. The tube is housed in a lead-lined aluminium jacket and the whole can be conveniently mounted on a stand. Left: the small high-tension unit for supplying the necessary 25 kV, D.C.

Owing to these favourable properties, the over-all blurring effect produced by the image intensifier is not more than 0.4 mm. In order that the advantage of this low "intrinsic" blurring may not be lost, the previously mentioned geometrical blurring must also be limited as much as possible. It is therefore desirable to use in conjunction with the intensifier an X-ray tube having a very small focus (fine focus tube).

Contrast

On passing through the subject to be examined, the X-rays are attenuated to an extent that varies from one point to another, and it is the resultant "radiation contrast" that supplies the required information regarding the subject, i.e. the image. The processes involved in the conversion of the X-rays to fluorescent light, then into photo-electrons and finally into light again, are wholly linear, owing to the very small electronic current occurring in the image intensifier. In principle, therefore, the contrasts in the image are unchanged (gamma = 1), in contradistinction to, for example, photographic reproduction (in which case gamma may be as much as 2.5 to 3).

For various reasons, however, the image intensifier does occasion a certain amount of fogging, i.e. a brightness contribution more or less uniformly distributed over the whole image, which results in a loss of contrast. The glass window of the tube through which the X-rays enter partly scatters the rays and thus delivers diffuse radiation to the fluoroscopic screen; similarly those X-rays which are not absorbed by the fluoroscopic screen but which fall on the wall of the tube are in part rediffused towards the screen. Again, a fraction of the blue fluorescent light which the very thin photocathode is not capable of absorbing is reflected from the inner wall of the tube, back to the photocathode, where it liberates electrons in the wrong places; up to a point this also applies to the persistence of the fluoroscopic screen. Furthermore, the photocathode exhibits a certain amount of thermionic emission which is independent of the intensity of illumination received by it; field-emission is also liable to occur at the photocathode. Both these effects contribute additional light uniformly spread over the whole of the viewing screen. A proportion of the total number of electrons arriving at the viewing screen is reflected there and tends to liberate positive ions from the walls of the anode space (as well as from possible residual gases), and these ions may be diffused in the direction of the anode aperture where they are accelerated by the field in the direction of the photocathode; here they may liberate more electrons and this will also produce additional diffuse light. Finally, at the viewing screen there is also some diffusion of the light generated by the screen itself, resulting in a kind of halo which is intensified by repeated reflection in the glass support of the screen, whose granules are in contact with it.

Very severe fogging would occur if a part of the light from the viewing screen were to return to the photocathode (a kind of "feedback"), but this is prevented by coating the rear of the viewing screen with a thin layer of aluminium which transmits practically all the high-velocity electrons, but completely intercepts light that would be reflected back [10]).

Suitable precautions have also been taken to reduce to a minimum all the other sources of fogging summarised above. As a result the loss in contrast is kept within reasonable limits. We will mention only one or two of these precautions. For example, the inner lining of the tube is of a kind that does not reflect blue light too well. The shape of the electrodes is such that high field strengths do not occur at the photocathode, and field emission is accordingly low. For the same purpose and also in order to limit thermionic emission, the threshold potential of the cathode should be fairly high.

Accurate data regarding the degree of fogging cannot be given at this time, but for orientation purposes we may assume that this represents 10% of the total luminous flux which the viewing screen would emit without the fogging. (This value is probably not very far from the actual figure.) From the aspect of contrast this means that where we would have had to observe contrast between a brightness of 105 and an average of 100 for example, we will now have a contrast of 115 : 110. The loss is thus of little importance. When contrast is very pronounced the depreciation is greater (10 as against 100 would become 20 to 110), but such contrasts are generally easily distinguished in any case.

Principle limitation of minimum perceptible subject detail

Let us now ascertain in how far the radiologist, using the image intensifier, really will be able to perceive more detail, that is, details of smaller dimensions and lower contrasts, than with the ordinary fluoroscopic screen. The ultimate answer to this question can be obtained only with the intensifier in actual use. While, however, insufficient medical experience has been gained as yet with the instrument to enable one to express in numerical values the advance which the introduction of the intensifier undoubtedly represents, it is possible to show theoretically that the minimum detail to be perceived cannot decrease below a certain limit. This is attributable to the circumstance that X-rays are in the form of quanta; attention was first drawn to this consequence by the scientists Sturm and Morgan [11]).

The energy quanta, or photons, of magnitude

---

[10]) See J. de Gier, Philips Techn. Rev. **10**, 97-104, 1948.

[11]) R. E. Sturm and R. H. Morgan, Screen intensification systems and their limitation, Amer. J. Röntgenology Rad. Ther. **62**, 617-634, 1949.

$h\nu$, of which all radiations consist (h = Planck's constant and ν = frequency of oscillation), are not emitted by the source perfectly regularly, nor are they absorbed regularly by a medium such as the fluoroscopic screen or the retina of the eye. It is possible to speak of an average number N of quanta per given interval of time, on a given surface, and this determines the intensity of the radiation emitted or absorbed. In the case of a number of such time intervals, however, the number of quanta within each interval will not be equal to N; random fluctuations about this average occur, whose order of magnitude, computed on the basis of probability, is given by the so-called standard deviation $\sqrt{N}$. The order of magnitude of the relative fluctuations in the radiation is defined by the coefficient of variation, which is equal to $\sqrt{N}/N = 1/\sqrt{N}$.

These natural fluctuations, the "noise" in the radiation, set a fundamental limitation on the detail which the radiation is capable of conveying, or, in more concrete terms, on the minimum radiation contrasts which can be communicated by means of the quanta absorbed.

This argument is immediately applicable to the human eye. Let us take as starting point a surface element of a given object, with a size of 0.1 mm^2. Those quanta will contribute towards the perception of brightness which travel from this surface element to the pupil of the eye within the "accumulation time" T of the eye. Only a fraction of these is effectively absorbed in the retina, i.e. produce the stimulus conveyed to the brain by the optic nerve. The fluctuations in this number of quanta can be regarded as fluctuations, in effect, in the observed brightness of the surface in question, and it is found that the difference in brightness between this surface and its background definitely cannot be perceived if the contrast is not at least 3 to 5 times as great as the relative fluctuation already mentioned. The lower limit of perceptible contrast is thus expressed in terms of the "noise" in the light quanta effectively absorbed. It will now be seen at once that as the size of the surface observed is increased (i.e. more quanta are permitted to contribute to the sensation of brightness), so also does it become possible to appreciate lower contrasts; see for instance the well-known contrast-detail diagrams obtained in the fluoroscopic examination of phantoms (artificial test-objects), *fig. 5* [12]). It will also be clear that when the brightness level is raised, contrast sensitivity increases. However, this takes place more slowly than might be concluded from the decrease in the relative fluctuations, since, when the brightness increases, the percentage of light quanta effectively absorbed drops (owing amongst other things to a saturation effect in the optic nerve), as also does the accumulation time T.

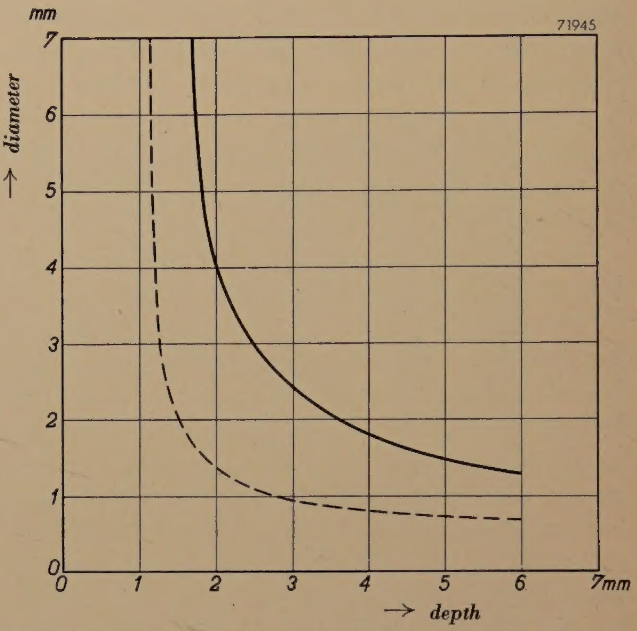


Fig. 5. Example of a contrast-detail diagram as obtained with the aid of an X-ray "phantom". The "phantom" consists of a number of plates of "Philite", one of which has in it a number of holes of different diameters and depths. The full line represents as a function of the hole diameter, the depth of a hole (contrast!) necessary to make it just visible in normal fluoroscopy. (The dotted line shows the results of provisional tests with the image intensifier under the same conditions.)

Let us now turn to X-ray diagnostics, again taking a surface of 0.1 sq. mm of the fluoroscopic screen. In chest examinations 100 000 X-ray quanta fall on such an area per second. Of this the fluoroscopic screen absorbs 65%. The accumulation time T of the eye in normal fluoroscopy is about 0.1 sec. The number of X-ray quanta involved in providing information about details of the above-mentioned size of the lungs under examination is therefore about 6500. One X-ray quantum of 70 kV absorbed liberates an average of 5000 light quanta in the screen. Thus we have more than 3×10^7 effective quanta; these occur in "packets" of 5000 quanta at a time, however. Consequently — and this is the cardinal point of the whole argument — the large number of quanta that would yield the low relative fluctuation of 1/5600 (if all the quanta occurred individually) cannot furnish any more information than the original 6500 X-ray quanta were capable of doing, with the much greater fluctuation of 1/80.

---

[12]) See G. C. E. Burger, Phantom tests with X-rays, Philips Techn. Rev. 11, 296, 1950.

In normal fluoroscopy this is not really important. In fact, of the light quanta obtained, only a very small proportion falls on the pupil of the eye (see *fig. 6*): in the retina only 120 quanta are effectively absorbed, which is much less than the available number of X-ray quanta. In X-ray images of objects 0.1 sq. mm in size, therefore, the eye will not perceive contrasts of less than 3 to 5 times 1/11 and a large part of the information contained in the X-ray quanta is thus not used.

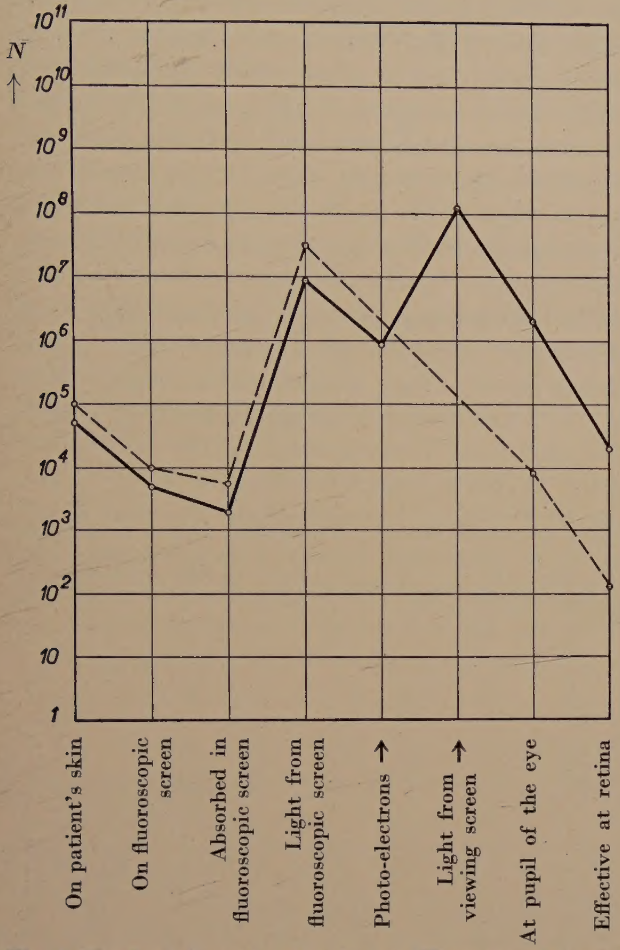


Fig. 6. Number of quanta occurring at different stages during a chest examination by means of X-rays, within a beam corresponding to an area of 0.1 sq.mm on the fluoroscopic screen. Voltage on X-ray tube 70 kV, current 2.5 mA, distance from X-ray focus to screen 70 cm.

Dotted line: ordinary fluoroscopy; accumulation time 0.1 sec.

Full line: with image intensifier; the accumulation time is now slightly shorter.

The image intensifier, however, has entirely altered this state of affairs. Initially the situation is rather less favourable, since, out of the above mentioned 100 000 X-ray quanta falling per second on 0.1 sq. mm of the thinner fluoroscopic screen, fewer are absorbed by the thinner screen and, moreover, the accumulation time T of the eye is shorter at the higher brightness level of the intensifier. We may take the number of effectively absorbed quanta to be about 1750. The light quanta produced by them in the fluoroscopic screen, however, are now put to better use (see fig. 6); it may be estimated that the area observed now yields roughly 20 000 actively absorbed light quanta, which is considerably more than the number of active X-ray quanta. This means that contrast sensitivity is then high enough for the eye to register the whole of the information contained in the X-ray quanta — but, obviously, not more.

This bears a direct analogy with optical enlargements. For example, in the image produced by a microscope objective, the fine details of the object are resolved only up to a certain limit; the "information" regarding the object is limited by the resolving power of the objective (which is in turn determined by the lens aberrations and ultimately by the wavelength of the light). The resolving power of the eye is much lower and the image is therefore observed through a high-magnification eye-piece. Now the greatest effective magnification of the eye-piece is that at which the details as separated by the objective can be seen as separate by the eye. If the eye-piece is any stronger, the details are certainly seen at larger angles, but the wealth of detail is not increased, the information contained in the image being by then already exhausted. It may be said that the extra enlargement is "ineffective". Similarly it may be said of the image intensifier that the part of the brightness intensification which corresponds to the factor by which the number of effective light quanta in the eye exceeds the number of effective X-ray quanta in the fluoroscopic screen, may be regarded as ineffective.

The conclusions to be drawn from the above argument are as follows:

1) Of the brightness intensification factor of 1000 which our instrument yields, a maximum factor of about 20 is "effective" [13]; the remaining factor of 50 plays no part in reducing the smallest perceptible contrast, but is of great practical significance as it makes vision easier (no adaptation necessary).
2) Should it be found desirable to make use of the properties of the image intensifier to reduce the X-ray dosage to the patient (for the same duration of observation), this will be unavoidably

---

[13] This factor is greater than the ratio 1750/120 to be derived from the example in fig. 6 since, owing to the decrease in the accumulation time T, as also in the percentage of quanta producing stimuli in the optic nerve, a relatively higher brightness is required to ensure that the eye absorbs as many effective light quanta as there are effectively absorbed X-ray quanta.

accompanied by a loss of "information" which is, after all, determined by the number of X-ray quanta. If the same amount of information as that obtained from the ordinary fluoroscopic screen is considered adequate, the dosage for the purpose of a chest examination could, theoretically, be reduced at most to about 1/15th; this would still leave in hand a gain in brightness to a factor of about 60, which would represent a marked decrease in adaptation time. For thicker subjects the permissible reduction factor would be smaller.

The image intensifier and radiography

Similar considerations to those outlined above will make it clear that the use of the image intensifier can never provide the same high quality of image as that obtained in X-ray photography (radiography). The photograph owes its superiority to the fact that the accumulation time of the photographic emulsion is not limited, as is that of the human eye, added to which the emulsion retains what it has accumulated, the dosage time being thus no longer than the accumulation time. Unless a series of exposures is to be made, the dosage is therefore not usually dangerous and the intensity of the X-rays and/or the exposure time can be adjusted for each subject so as to yield the best average density. For a representative grade of film it will be found that to this end 20 000 X-ray quanta per 0.1 sq.mm (about 0.001 ergs of energy) must be effectively absorbed. The exposed film is examined with the aid of a viewing lantern; the illumination level can be raised as desired and matters so arranged that the light quanta received by the retina are in any case sufficient for perception of all the information the photograph has to offer (*fig.* 7). Now the difference in image quality between normal fluoroscopy, image intensifier and radiography is immediately expressed by the difference in number of useful quanta, i.e. for a chest examination 120 : 1750 : 20 000. With subjects that entail a higher degree of absorption (e.g. gastric fluoroscopy), the supremacy of the photograph is still more pronounced; it remains at 20 000 quanta, whereas in the two other cases the numbers of quanta are much lower.

Under these circumstances it will be obvious that the principle of the electronic image intensifier has no advantage over conventional X-ray photography. It can be employed to advantage, however, in fluorography. In this Review, on several occasions, the fact has been discussed that in this technique, whereby a reduced X-ray image is obtained with the aid of a camera, it is a difficult matter to obtain sufficient light to expose the film, and this is the reason why extremely fast lenses have been developed for this class of work [14]). The image intensifier, which furnishes a very bright, reduced image that is capable of being photographed quite easily on a 1 : 1 scale, provides a very convenient solution to the problem. Apart from this, the image intensifier may possibly prove to be the answer to the old problem of X-ray cinematography. As far as the mechanical aspect is concerned, X-ray fluorography has already brought a solution within sight, but the next step towards successful X-ray cinematography, although apparently only a small one, calls for some caution, since a very large number of photographs taken in succession means that the total X-ray dosage to the patient will again be fairly high. A reduction in the X-ray intensity with a view to limiting the total dosage

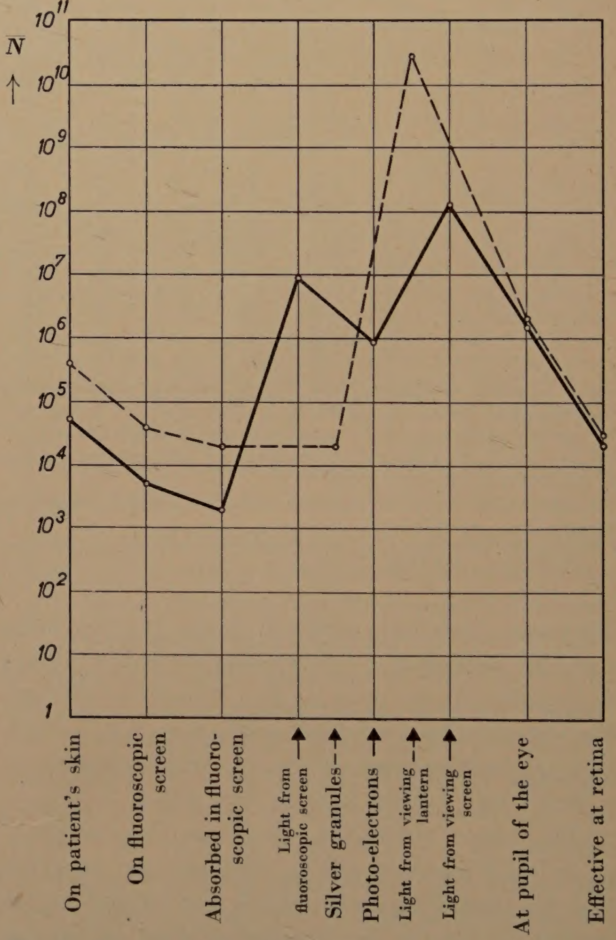


Fig. 7. Same as in fig. 6 but for normal radiography (dotted line). Full line; with image intensifier, as reproduced from fig. 6.

---

[14]) W. Hondius Boldingh, Fluorography with the aid of a mirror system, Philips Techn. Rev. **13**, 269-281, 1952 (No. 10).

during the taking of the film is possible only up to a certain point, seeing that even with very fast lenses too little light is otherwise obtained to ensure sufficient density of the film. This difficulty is overcome by the extra brightness provided by the image intensifier; even at appreciably lower X-ray intensities than those usually employed in fluorography, the image on the viewing screen will ensure a well-exposed film without necessitating particularly fast lenses. The fact that a reduction in the X-ray intensity detracts from the image quality (available information) has already been explained. It is a well-known phenomenon, however, that some slight sacrifice in image quality can safely be made in cinematography, since the summation of the impressions received from the successive pictures provides additional information.

Summary. In order to reduce to a minimum the risk of injury to the patient as a result of X-ray examination, the X-ray dosage that the patient receives must be limited as far as is practicable. Therefore, the attainable brightness of the image on a standard fluoroscopic screen is so low that long visual adaptation periods are required before such images can be studied; even then, visual acuity and the contrast sensitivity of the eye are both fairly low at such low brightness levels. Improvement can be brought by increasing the brightness of the fluoroscopic image ("intensifying"). The electronic tube by means of which this is now made possible functions on the principle of the image converter. The X-rays produce on a fluoroscopic screen an image whose radiation liberates electrons from a photocathode in contact with the screen. The electrons are focussed by an electrostatic electronic lens system working at a potential of 25 kV, to produce another image, reduced 9 times in size, on a second fluorescent screen, the viewing screen, which is observed through a simple microscope of 9 × magnification. The gain in brightness, to a factor of 800-1200 with respect to normal fluoroscopy, is produced partly by the accelerating field which imparts energy to the electrons, and partly by the electron-optical reduction of the fluoroscopic image. The resolving power of the image intensifier is higher than that of the best fluoroscopic screen. Owing to a number of causes the ultimate image is subject to a certain amount of fogging, which does reduce the contrast, although not to a serious degree. From a discussion of the lower limit of perceptible detail as based on spontaneous fluctuation in the number of X-ray and light quanta it is seen that the brightness can be intensified up to a factor of about 20 with a corresponding increase in the amount of "information" to be derived from the image. Beyond that point the information is restricted by the number of contributing X-ray quanta, but the much higher gain factor of the image intensifier offers the advantage that the image can be viewed without preliminary dark adaptation of the eyes, in a room with practically normal illumination. In conjunction with film cameras the intensifier paves the way to X-ray cinematography without the necessity for increasing the X-ray dosage to a hazardous extent.

MAGNETRONS

by J. VERWEEL.

621.385.16:621.396.615.141.2

Split-anode magnetrons have already been in use for a number of years for generating decimetric and centimetric waves. In 1937 a radio telephone link between Eindhoven and Nijmegen (Holland) employing magnetrons and operating on a wavelength of 25 cm was established. During the period 1940-1945 great strides were made in the development of magnetrons; output power was increased by more than 1000 times and generation of waves of 1 centimetre became a practical possibility. In this way the magnetron came to assume the main role in radar, and that this has had an important effect on war-time as well as peace-time activities is sufficiently well-known. In this article the working and construction of modern magnetrons are discussed.

Some 30 years ago **Hull** 1) published an article on the effect of an uniform axial magnetic field on the motion of electrons between coaxial cylinders (*fig. 1a*). Although **Hull** spoke neither of oscillations, nor used the word magnetron, his work laid the foundations for the use of the magnetron as an oscillator. Since 1921 a varying but none the less steadily increasing amount of interest was shown in the magnetron. Up to 1930 the oscillations produced by these tubes were of little practical value, but from that time onwards there was a widespread increase in interest, as it was found that split anodes (fig. 1*b*) with resonators or transmission lines between them were capable of generating frequencies with a high degree of efficiency. **Posthumus** 2) has contributed much towards the development of valves designed on this principle and his theory of the oscillations produced by them is regarded as of pioneer value. In the four years following his publication more than 100 articles were written on the subject of the magnetron, in all parts of the world.

In the period up to 1940 interest waned slightly, but in that year **Boot** and **Randall** 3) carried out experiments with magnetrons provided with cavity resonators (fig. 1*c*), and these experiments led to the successful use of the magnetron as a transmitting tube for radar. Feverish activity ensued; output power was increased once more and some very short waves were produced. Important theoretical work was also submitted. To give some idea of what has been achieved with these tubes, we may mention that in recent years magnetrons have been made which operate at a wavelength of 3 cm, delivering more than 1000 kW peak output, and that others operate at a wavelength of 3 mm, supplying an output of several kilowatts 4). This power is transmitted in pulses varying in duration from a few tenths of a microsecond to several microseconds, the repetition frequency being usually between 500 c/s and 4000 c/s. *Fig. 2* illustrates a modern magnetron for 8.5 cm waves, delivering a power of 400 kW peak.

By far the most important field of application of the magnetron to date is **radar**, after which comes **diathermy**, where centimetric waves are in some cases particularly beneficial. Another special field is to be found in the **linear accelerator**, an appliance for imparting enormous velocities to electrons for the purposes of nuclear physical research 5).

In 1939 this Review contained an outline of the development of magnetrons up to that time 6); the present article is intended to provide a general review of modern magnetrons which will at the same time pave the way for later articles dealing with special designs in this sphere. In the present paper we shall endeavour to answer the following questions.

1) In what manner is the energy derived by the electrons in the magnetron from the direct voltage source converted to H.F. energy?
2) To what conditions must the resonators conform?
3) What kind of output system can be used to draw from the magnetron the energy developed by it?

---

1) A. W. Hull, Phys. Rev. **18**, 31-57, 1921.

2) **K. Posthumus**, Principles underlying the generation of oscillations by means of a split anode magnetron, Philips Transm. News **1**, 11-25, Dec. 1934; Oscillations in a split anode magnetron, Wireless Eng. **12**, 126-132, 1935.

3) **H. A. H. Boot** and **J. T. Randall**, The cavity magnetron, J. Inst. Elec. Engrs. **93**, 3a, 928-938, 1946.

4) **J. R. Pierce**, Millimeter waves, Electronics **24**, 66-69, 1951 (No. 1)

5) Philips Techn. Rev. **14**, 1-12, 1952 (No. 1).

6) **G. Heller**, Philips Techn. Rev. **4**, 189-197, 1939.

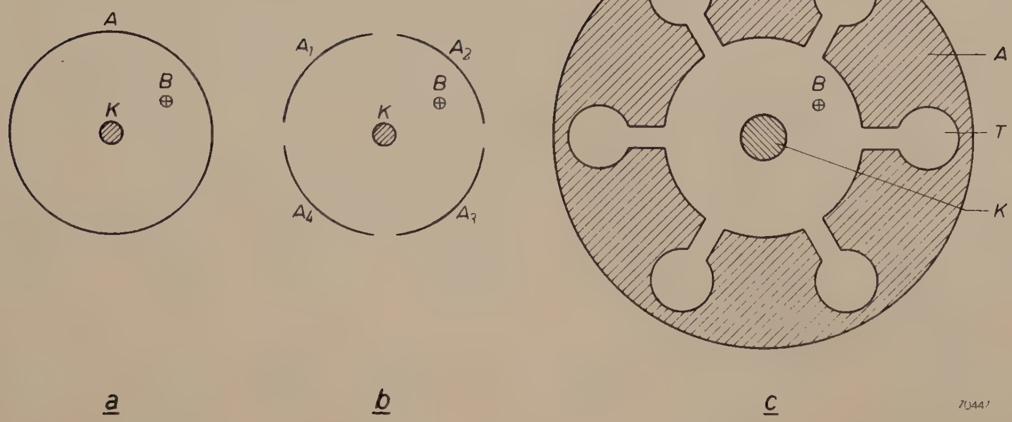


Fig. 1. Various forms of magnetrons in cross section. K = cathode; A = anode. The flux density B is perpendicular to the plane of the drawing and is uniform within the whole of the anode space. *a*) Hull magnetron. *b*) Magnetron with split anode, comprising four segments (A_1, A_2 A_2 and A_4). Such magnetrons can have a very high efficiency. *c*) Magnetron by Boot and Randall. The anode is a solid block of copper with cavity resonators T cut in it.

This will be followed by a few details regarding construction.

It will be necessary to restrict ourselves to the main essentials and to refer the reader to existing publications [7]) for a more detailed study.

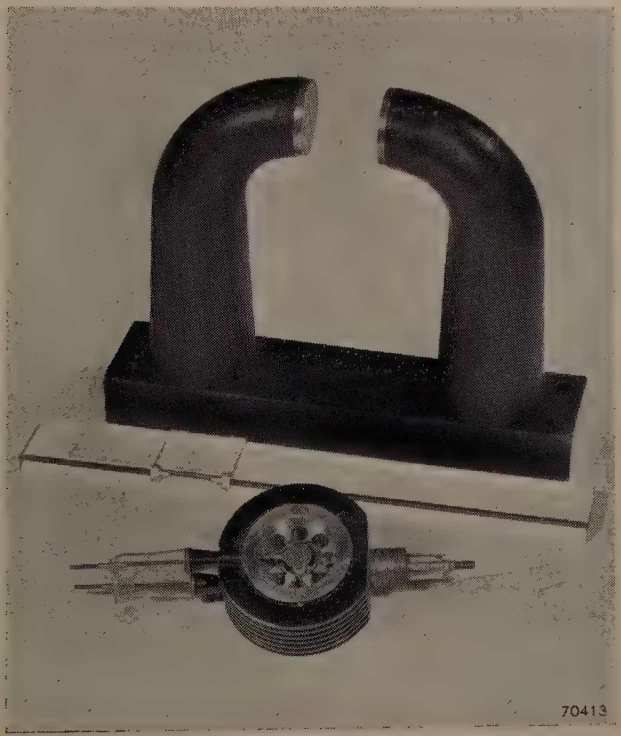


Fig. 2. Experimental magnetron for a wavelength of 8.5 cm, 400 kW peak output, with one of the covers removed. Behind the magnetron will be seen the associated permanent magnet.

Motion of an electron in mutually perpendicular electric and magnetic fields

In a cylindrical magnetron the electric field is mainly perpendicular to the magnetic field. When the magnetron is oscillating, the electric field has not only a radial, but also a tangential component and, moreover, the space charge of the electrons has a very marked effect on the motion; in consequence the theory is rather complicated. As the various computations that have been made are very lengthy, they fall outside the scope of this article and we shall therefore confine ourselves to a general outline of the electronic motion.

Let us first call to mind the theory relating to the motion of electrons in a uniform magnetic field, the flux density of which is denoted by B (in the z-direction of a rectangular system of co-ordinates), and a uniform electric field E perpendicular thereto (in the negative y-direction). We are then concerned with the electric and the Lorentz forces. Analytically this is represented in our system of co-ordinates by:

$$\begin{aligned} m\ \mathrm{d}v_x/\mathrm{d}t &= ev_yB,\\ m\ \mathrm{d}v_y/\mathrm{d}t &= eE - ev_xB,\\ m\ \mathrm{d}v_z/dt &= 0, \end{aligned}$$

where m is the mass of the electron, t is the time and v_x, v_y and v_z are the velocity components along the axis.

It is easy to verify the general solution of these expressions, viz:

$$v_x = -a\cos\frac{eB}{m}(t-t_0) + \frac{E}{B},$$

---

[7]) See e.g. G. B. Collins, Microwave magnetrons, Radiation Lab. Series 6, McGraw Hill, New York 1947, and J. B. Fisk, H. D. Hagstrum and P. L. Hartman, The magnetron as a generator of centimeter waves, Bell Syst. Techn. J. **25**, 167-348, 1946.

$$v_y = a \sin \frac{eB}{m}(t - t_0),$$

$$v_z = b.$$

a, b and t_0 are integration constants. Further integration gives us the equations for the electron paths. If we assume that all co-ordinates and velocities are zero at $t = 0$, we have:

$$\left.\begin{aligned} x &= \frac{E}{B}t - \frac{m}{e}\frac{E}{B^2}\sin\frac{eB}{m}t, \\ y &= \frac{m}{e}\frac{E}{B^2}\left(1 - \cos\frac{eB}{m}t\right), \\ z &= 0. \end{aligned}\right\} \quad \dots \quad (1)$$

These equations represent a cycloid, as described by a point on the periphery of a circle rolling along the x-axis with constant velocity E/B (*fig. 3*). The diameter of the circle is $(^{2m}/_e)$ $(^E/_{B^2})$, and this represents the maximum distance of the electron from the x-axis. The general movement of the electron is perpendicular to the lines of electric and magnetic force (i.e. in an equipotential surface of the electric field) and the electron thus describes arcs to a height of $(^{2m}/_e)$ $(^E/_{B^2})$.

Now let us take the case of a plane magnetron, the cathode of which is formed by the plane $y = 0$ and the anode by $y = d$ (*fig. 4*). As long as the magnetron does not oscillate and, disregarding the space charge, the field between the cathode the anode will indeed be uniform. If, as assumed, the electrons leave the cathode with zero velocity, they will move along the cycloidal paths mentioned previously. If $(^{2m}/_e)$ $(^E/_{B^2}) > d$, they describe only the first part of the cycloid shown in fig. 3 and are then absorbed by the anode. As soon as $(^{2m}/_e)$ $(^E/_{B^2}) < d$, however, the cycloidal path is completed and the electrons do not reach the anode. The magnetic field whereby the path just touches the anode is known as the critical field, the flux density of which is denoted by B_{cr}. Hence:

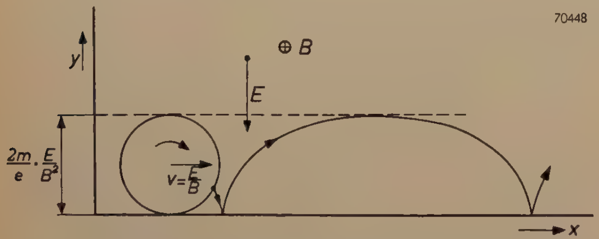


Fig. 3. Under the combined influence of a uniform electric field E and a perpendicular uniform magnetic field (flux density B), an electron emitted with zero velocity at the x-axis follows a cycloidal path. The average velocity in the x-direction is E/B.

$$B_{cr}^2 = \frac{2m}{e}\frac{E}{d},$$

$$B_{cr} = \sqrt{\frac{2m}{e}\frac{\sqrt{V}}{d}},$$

where V is the anode voltage.

For a cylindrical magnetron, in which perfectly cycloidal paths are obviously no longer possible, a similar equation applies, viz:

$$B_{cr} = \sqrt{\frac{8m}{e} \cdot \frac{\sqrt{V}}{r_a\left(1 - r_k^2/r_a^2\right)}}, \quad \dots \quad (2)$$

where r_a is the radius of the anode and r_k that of the cathode. In a non-oscillating magnetron the anode current is therefore zero when $B > B_{cr}$.

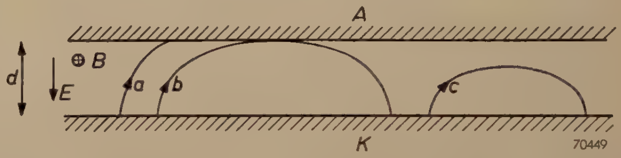


Fig. 4. Electron paths in a non-oscillating magnetron with distance d between cathode K and anode A (both plane):
a) $(^{2m}/_e)$ $(^E/_{B^2}) > d$: paths end at the anode;
b) $(^{2m}/_e)$ $(^E/_{B^2}) = d$: paths touch the anode;
c) $(^{2m}/_e)$ $(^E/_{B^2}) < d$: electrons no longer able to reach the anode.

When the electric field is no longer uniform, but nevertheless still perpendicular to the magnetic field, as in an oscillating magnetron, the motion is more complex. Assuming the magnetic field to be strong enough, so that the radius of curvature of the cycloidal path is small compared with that of the equipotential lines, the electric field in small areas can be regarded as uniform. In that case, too, the electron, on an average, follows an equipotential line. It is true that in the actual magnetron the radius of the circular motion is not usually small in comparison with that of the equipotential lines, but the approximation is close enough to justify a qualitative representation of the mechanism.

Qualitative explanation of the occurrence of oscillations

In the event that $B < B_{cr}$, the magnetron behaves as an ordinary diode and oscillation is not possible. It will therefore be investigated what happens when $B > B_{cr}$. In the absence of oscillation no anode current flows and the electrons follow cycloidal paths, their general movement being parallel to the cathode and anode. When oscillation occurs the conditions are different; the electrons are then able to follow paths which do terminate at the anode; anode current flows and the magnetron consumes energy. Let us now assume that the condition of oscillation has been established in one way or another, and investigate how such oscillations are sustained.

Let the number of anode segments be $2n$ and assume the potentials between two adjacent segments to be constantly varying in counterphase, as is usually the case. *Fig. 5* shows the lines of electric force of the alternating field at a given moment, in the event of four segments. According to Posthumus, this stationary alternating field is equivalent to fields rotating in opposite directions; the two fundamental components of these fields rotate with an angular velocity ω/n or $-\omega/n$, where ω is the angular frequency of the oscillation. Now, it is essential to the mechanism that the electrons be synchronised with one of these fields. The mean tangential velocity of the electrons, as we have seen, is E/B, An approximate expression for the synchronisation can be based on the velocity of the electrons half-way between anode and cathode, disregarding the fact that the electric field is dependent on the radius r. The angular velocity is then:

$$\frac{v}{\frac{1}{2}(r_a+r_k)}=\frac{E/B}{\frac{1}{2}(r_a+r_k)}=\frac{V/(r_a-r_k)B}{\frac{1}{2}(r_a+r_k)}=\frac{2V}{(r_a^2-r_k^2)B}\,.$$

Equating this with the angular velocity ω/n, we then have:

$$V=\frac{\omega}{2n}r_a^2B\left\{1-\left(\frac{r_k}{r_a}\right)^2\right\}\quad .\ .\ .\ .\quad (3)$$

We shall refer to this equation again presently, but assume for the moment that it is satisfied by a suitable selection of the parameters.

Let us now examine the interaction between the electron and the synchronous field, assuming that both rotate in, say, a clockwise direction. The other rotating field passes the electrons very quickly; as this is an alternating field it will have alternately an accelerating and a decelerating effect. The electrons will, therefore, perform small oscillations about the average path, but these will be disregarded. Let us first see how the electron behaves in an area where the tangential field has a decelerating effect (at A in *fig. 6*). As the condition of synchronisation has been met, the electron will remain in this zone. It is true that the alternating field is usually smaller than the steady field, but the effect may nevertheless not be ignored.

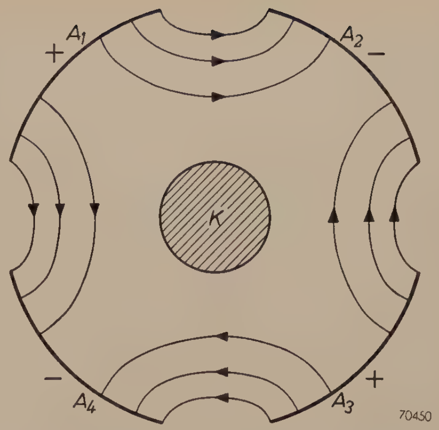


Fig. 5. The alternating electric field in a magnetron with four plates is a standing wave in the space between anode and cathode. This wave may be resolved into opposed rotating waves, or rotating fields. The electrons interact mainly with one of these fields.

The resultant E of the two thus forms a certain angle with the radial direction and, since the average motion of the electron is in the direction perpendicular to E, it will be seen that it is deflected towards, and ultimately reaches, the anode. Disregarding the fact that the electric field strength is dependent on the radius, the average velocity E/B of the electron, and therefore also the kinetic energy, will be constant. The potential energy dissipated by the electron in passing to the anode is thus not converted to kinetic energy as in the case of normal motion in an electric field, but is imparted to the alternating field as H. F. energy.

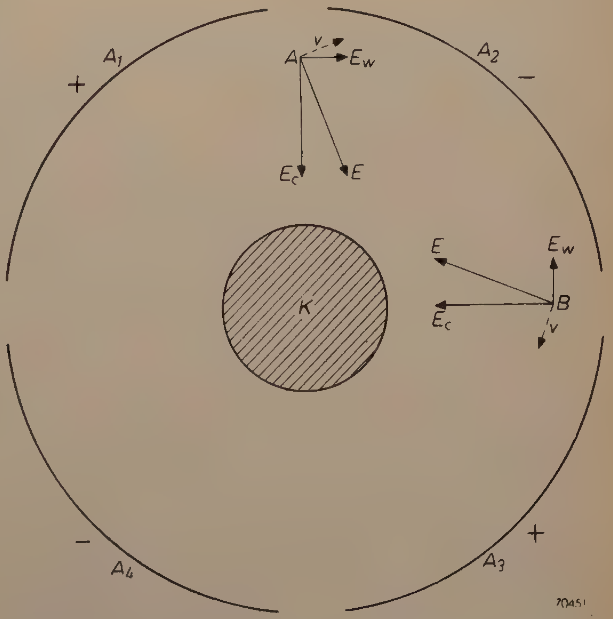


Fig. 6. The resultant E of the steady field E_c and the alternating field E_w, at two points, A and B. The mean electron velocity v is perpendicular to E. In a decelerating field E_w, v receives an outward-directed component, in an accelerating field E_w, an inward-directed component.

An electron in an area where a tangential **accelerating** field prevails (at B in fig. 6) will follow a path towards the cathode, since the field vector points in the other direction. The electron is accelerated by the H.F. field and is thrown back to the cathode after describing one arc of the cycloid. The paths of two such electrons are depicted in *fig. 7;* it will be seen that the first of the two describes a number of arcs (a), but that the second completes less than a single arc (b). The time required to describe one complete arc is constant and is, according to equation (1), equal to $(^{m}/_{eB}) \cdot 2\pi$. Seeing that the electrons a are in the field longer than the electrons b and therefore cede more energy than is absorbed by the electrons b, there remains a surplus, so that oscillation can be sustained and energy can be delivered to an external system.

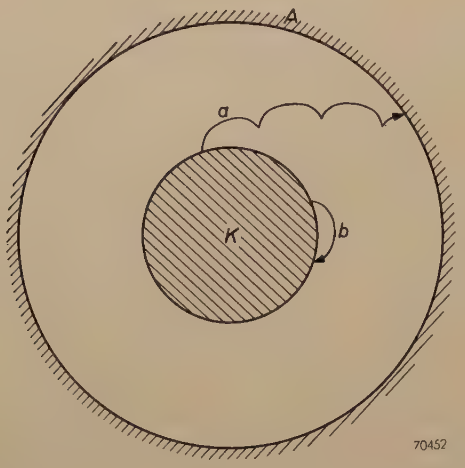


Fig. 7. Interaction of an electron with (a) a decelerating field and (b) an accelerating field. Electrons (a) are in the field longer than electrons (b), so that, averaging over all the electrons, energy is supplied to the field.

It is, furthermore, a fortunate circumstance that electrons in a magnetron are driven towards a zone where a field of maximum deceleration exists and where the energy transfer is as favourable as possible; in effect, **phase focusing** occurs, as will be seen from *fig. 8*, in which the resultant E of the steady field E_c and the alternating field E_w are shown at the points A as well as at the points C and D, on either side of it. An electron situated at C for example, because it lags, will be in a stronger field than an electron at A. The mean velocity E/B of such an electron will thus be higher until the other field is overtaken. At D the electrons are in a weaker over-all field E; they therefore move more slowly until the field catches up with them. In consequence of this, electron bunches occur at the most advantageous point and impart H.F. energy to the field.

At this juncture we may point out the difference that exists between a magnetron, and an "inverted" cyclotron. In a cyclotron the charged particles move under the influence of a magnetic field and are accelerated twice per revolution by a tangential electric field whose H.F. oscillations are synchronised with the circular motion. It would be possible to imagine this process inverted, that is to say, a process of shooting high-velocity electrons inwards; they would then yield energy to an H.F. field at the slots and would thus be decelerated. This is certainly possible in the magnetron, but the mechanism is rather different from that just described; if the oscillations were promoted by means of a suitable combination of potential and magnetic field, this would result in oscillations of very low efficiency. The angular frequency of the oscillations corresponds to that with which the cycloidal arcs are traversed in the magnetron: $\omega = eB/m$. In the oscillatory mechanism of the magnetron this frequency is of minor importance, however; in this case it is the mean velocity E/B that is important, where E contains particularly the constant-field component, which is entirely absent in the cyclotron.

The electronic motion can be computed quantitatively by iteration; the electron paths are calculated numerically for given values of the direct and alternating anode voltages and of the magnetic field. From a sufficiently large number of paths the space-charge distribution in the magnetron is then obtained, and from this follows a new potential field, which, in general, is different from the original. Further electron paths are next computed from this new field; these in turn give rise to a third field and so on. With a judicious choice of the original field, these successive approximations converge upon the correct solution. In 1928 **Hartree** [8]) employed this method to quantum-mechanical path computations in atomic theory and, during the last war, the electronic paths in magnetrons

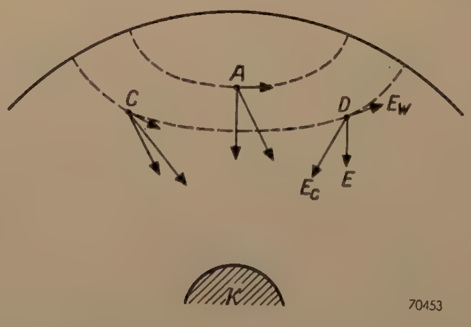


Fig. 8. The resultant field vector E at a point A, and at two other points C and D on either side of A. Owing to the influence of the radial alternating field, E is greater at C and smaller at D than at A.

---

[8]) D. R. **Hartree**, Proc. Cambridge Phil. Soc. **24**, 111-132-1928.

were computed in this manner under his guidance[9]).

The result of such an iteration procedure as applied to a magnetron having eight anode segments is seen in *fig. 9*. The paths of four electrons emitted by the cathode in different phases of the H.F. oscillation are depicted in a system of co-ordinates which rotates with the field. An electron *a* emitted at an unfavourable instant returns to the cathode; the three other, favourably emitted electrons *b*, *c* and *d* proceed to the anode and, in so doing describe a number of loops. Since the system of co-ordinates is rotating, the field is constant with time; hence every electron emerging from the cathode surface at a point which rotates with the co-ordinates, describes the same path within the system. The envelope embracing all these paths, shown by the broken line in fig. 9, represents the boundary of the space-charge cloud. This consists of four "spokes", each of which lies roughly at the location of the maximum deceleration field. They rotate together with an angular velocity $\omega/4$, and constantly pass those parts of the anode that are just about to become positive, intensifying the charge by reason of the induction. In this way the oscillation is sustained in much the same way as that of a pendulum to which a pulse is applied each time at the correct moment. It is thus clearly seen that the potential energy dissipated by the electron and not converted to kinetic energy, becomes available as H.F. energy.

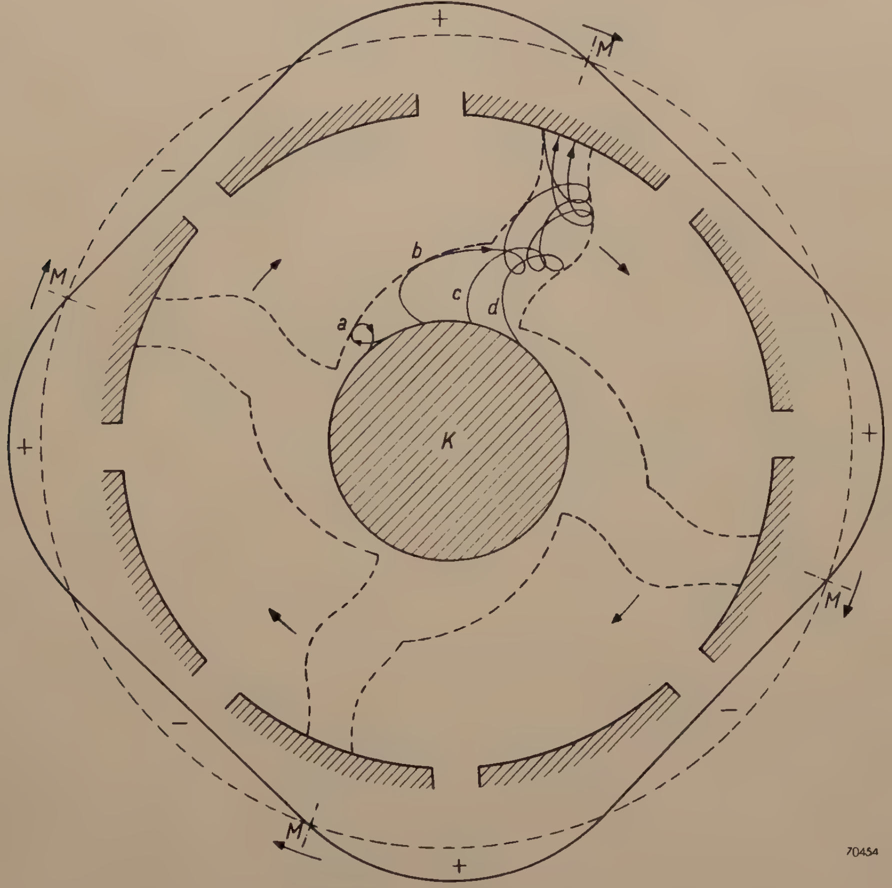


Fig. 9. Paths of four electrons (*a*, *b*, *c* and *d*) emitted by the cathode *K* in different phases with respect to the rotating field, plotted in a system of co-ordinates that rotates with the field. The actual field is constant with time. From a point on the cathode which rotates with the system of co-ordinates, the same path is traced each time. The envelope of all the paths is shown by the broken lines; this is the boundary of the space charge. The areas of the tangential field of maximum deceleration are shown by the lines *M*. The radial distance between the dotted circle and the curve fluctuating around it represents the rotating alternating field. (From an article by J. B. Fisk, H. D. Hagstrum and P. C. Hartman, Bell Syst. Techn. J. **25**, 195, 1946).

---

[9]) C. V. D. Report, Mag. 36 and 41.

Characteristics of the magnetron

In this section we shall trace the relationships that exist between current, voltage, magnetic field, flux density, etc.

We have already mentioned the cut-off condition indicating the smallest flux density necessary for oscillation, viz:

$$B_{cr} = \sqrt{\frac{8m}{e}} \frac{\sqrt{V}}{r_a(1-r_k^2/r_a^2)} \quad . \; . \; . \; . \quad (2)$$

Further, we have derived the approximate condition for synchronism.

$$V = \frac{\omega}{2n} r_a^2 B \left\{ 1 - \left(\frac{r_k}{r_a}\right)^2 \right\} . \quad . \; . \quad (3)$$

A more accurately defined condition for synchronism (which will not be discussed here) is as follows:

$$V = \frac{\omega}{2n} r_a^2 B \left\{ 1 - \left(\frac{r_k}{r_a}\right)^2 \right\} - \frac{m}{2e} \left(\frac{\omega r_a}{n}\right)^2 \left\{ 1 - \left(\frac{r_k}{r_a}\right)^2 \right\} . \quad (4)$$

Hartree formulated a condition that has to be met if the electrons are to reach the anode in the case of an infinitely small H.F. field. This states that:

$$V \geq \frac{\omega}{2n} r_a^2 B \left\{ 1 - \left(\frac{r_k}{r_a}\right)^2 \right\} - \frac{m}{2e} \left(\frac{\omega r_a}{n}\right)^2 . \; . \; . \quad (5)$$

In *fig. 10* all these lines are shown diagrammatically, with the flux density B as abscissa and the voltage V as ordinate. The cut-off curve (*2*) is a parabola; the "simple" condition for synchronism (*3*) is a straight line through the origin, and Hartree's condition (*5*) produces another straight line, parallel to the other and tangent to the parabola. The more accurate condition for synchronism (*4*) is represented by a line located between the other two straight lines and also lying parallel to them.

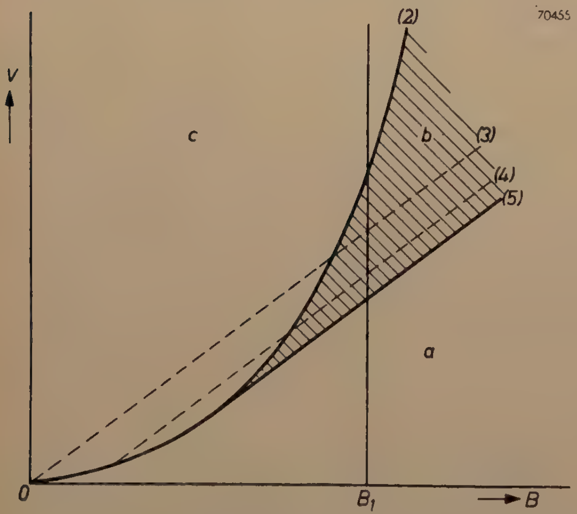


Fig. 10. The curve (*2*) representing the critical flux density and the Hartree line (*5*) divide the V-B plane into three zones *a*, *b* and *c*. In (*a*) the magnetron passes no current and no oscillation occurs. (*b*) is the oscillation zone; here the current rises steeply with the voltage. In (*c*) the magnetron is no longer cut off so that no oscillation occurs.

Lines (*3*) and (*4*) represent the simple and the more accurate conditions for synchronism respectively.

For the equations of lines (*2*), (*3*), (*4*) and (*5*) see the formulae of the same numbers.

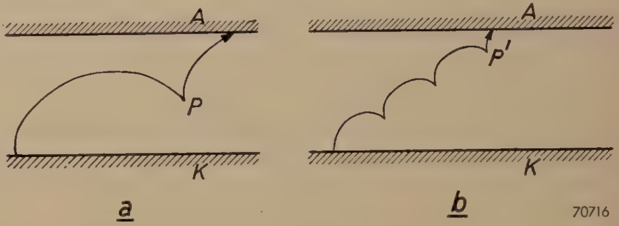


Fig. 11. Electron paths in a plane magnetron with cathode *K* and anode *A*.
a) *B* slightly higher than B_{cr}. The electron describes a large arc. As its velocity is zero at *P*, and as it can cede only little energy to the field on its way to the anode, a large proportion of the potential energy which it possesses at *P* is converted to heat. The efficiency is therefore only low.
b) *B* much greater than B_{cr}. The electron describes a small arc. Only that potential energy which the electron still possesses at *P′* can be converted to heat. The efficiency is now high.

Lines (*2*) and (*5*) divide the V-B plane into three zones *a*, *b* and *c*; if the voltage be increased for a given value B_1 of the flux density, each of these zones will be traversed in turn. At low potentials zone *a* is involved; the magnetron is then cut off, but oscillation is not possible, for the voltage does not conform to condition (*5*). When the voltage is raised so that zone *b* is entered, oscillation can occur. With further increases in potential the anode current rises rapidly and, with it, the amplitude of the oscillation. Usually a limit is then encountered owing to the fact that the cathode is unable to deliver more current. The oscillations, which are at a maximum in the region of line (*4*), decrease when the voltage is raised, seeing that the condition for synchronism is then not met. In zone *c* current may flow without H.F. voltage, the magnetron is no longer cut off and no further oscillation occurs.

The efficiency of the oscillation is found to rise when the flux density is increased (hence with increasing voltage in order to remain on line (*4*) which represents the condition for synchronism). This will be readily appreciated from the figure. If the flux density *B* is but slightly higher than the critical value B_{cr}, the conditions will be as shown in *fig. 11a*. As is seen from equation (*1*), the height of the arc of the cycloid is not much less than the distance from anode to cathode. An electron leaving point P roughly at zero velocity still possesses much of its potential energy and can impart only a small proportion of it to the H.F. field; a large proportion is thus dissipated on collision with the anode. The efficiency is then relatively low.

However, if B is much greater than B_{cr}, we see from fig. 11b that the electron cannot lose more potential energy than it still possesses at the point P' and as this is much less, less is dissipated and the efficiency is higher.

A magnetron for a wavelength of 3 cm may in this way yield efficiencies of 50%, which is much higher than is the case with klystrons or travelling-wave tubes. Electron bunching takes place in magnetrons in the same way as in these other types of tube and, moreover, the bunching is synchronised with the wave in magnetrons as in travelling-wave tubes. The difference, however, is that in the magnetron the magnetic field has a phase focusing effect. As with an ordinary top, a force in the one direction, in this case tangential, is converted to a displacement in the perpendicular direction, i.e. radial. The velocity in the direction of the field remains the same; there is no dissolution of the bunches, but the potential energy lost by the electron on its way to the anode is imparted, in the form of H.F. energy, to the alternating field, as already stated above. Accordingly, almost the whole of the D.C. energy is effectively used.

The characteristics which are usually plotted for magnetrons are combined to form what is known as a performance chart; the anode current is plotted horizontally and the anode voltage vertically. By varying the voltage and the magnetic field the various points in the diagram are obtained. The output power and efficiency are measured, and lines of constant flux density, power and efficiency are drawn in after the manner of *fig. 12*. The different phenomena which we have just described then become apparent. The anode voltage is roughly proportional to the flux density and depends but little on the current strength. Efficiency is increased with the flux density, as already pointed out. Further, the efficiency is to a certain extent dependent on the anode current; with high values of the current, efficiency is impaired, since the space-charge then disturbs the phase focusing.

In the direction of higher voltage and current the performance of the magnetron is restricted by the cathode emission, disruptive discharges in the tube and by the maximum magnetizing force which the magnet is capable of producing. For normal use, magnetrons are equipped with a permanent magnet (fig. 2), for which reason only those points in the performance chart can be attained which lie on a line of constant flux density.

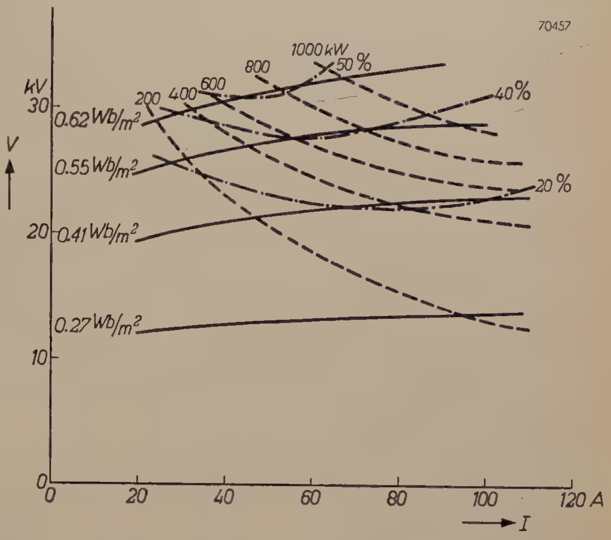


Fig. 12. Performance chart of a magnetron for a wavelength of 3 cm. The current is plotted horizontally and the voltage vertically. Lines of constant flux density (expressed in Wb/m^2), constant power (in kW) and constant efficiency (%) are shown in the chart. This particular chart was prepared from a magnetron having 18 cavities; r_k = 0.32 cm; r_a = 0.52 cm; anode length = 2 cm.

Resonators

The H.F. energy occurs mainly in the resonators, where it oscillates in the form of electric and magnetic energy. The resonators determine the wavelength and also serve as reservoirs for the energy. A portion of the energy in the resonator is fed to the aerial or to another form of load.

Prior to 1940 the preferred form of resonator was a two-wire transmission line. In two-plate magnetrons the two anode sections were connected direct to the two wires (*fig. 13a*), whereas in magnetrons with four or more plates connection was established in pairs by means of the shortest possible leads, to keep the pairs at the same voltage; the transmission line was then connected to these leads (fig. 13b).

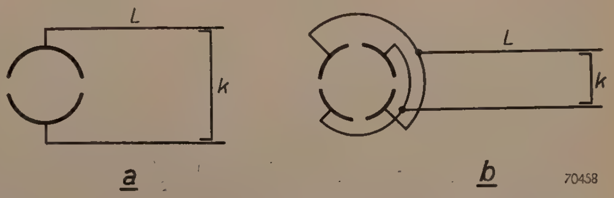


Fig. 13. a) Two-plate magnetron with transmission line L. Tuning is effected by means of the shorting bridge k.
b) Four-plate magnetron with plates connected in pairs. The transmission line is welded to these connections.

Present-day magnetrons usually work with several resonators in the form of cavity resonators, cut out of a solid block of copper which serves as anode. The dissipative and radiation losses in such resonators are smaller, the dissipation of heat is greatly simplified and the resonators can be made more easily with a sufficiently high degree of accuracy for use on short wavelengths. Conventional forms of magnetron anode are depicted in *fig. 14*.

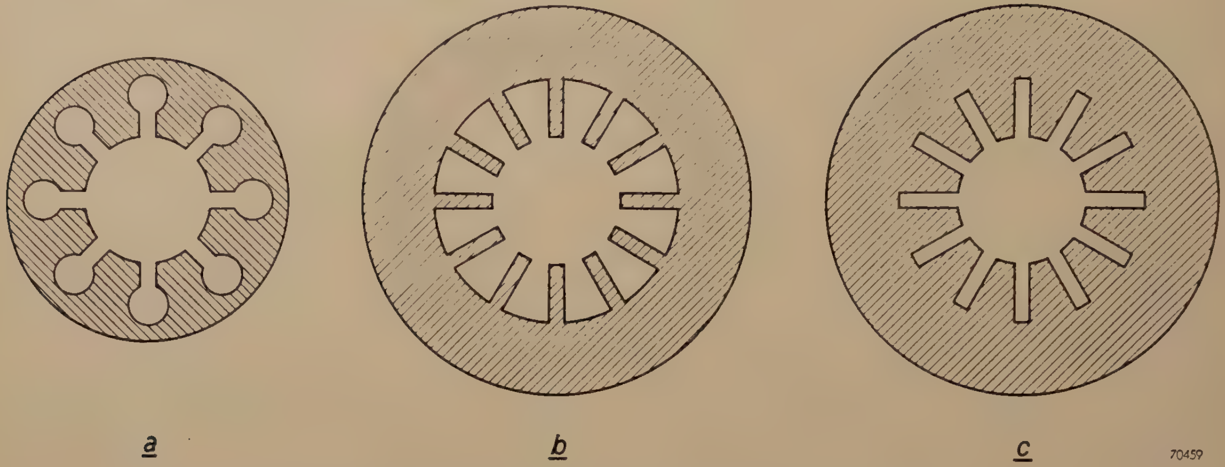


Fig. 14. Anode blocks for magnetrons (shown in cross section). The cavities consist of (*a*) holes and slots; (*b*) sectors; (*c*) slots.

In the case of resonators consisting of holes and slots it is still permissible to speak of lumped inductances and capacitances. The capacitance is concentrated particularly in the slots and we may also refer to a certain potential across the slots. The inductance may be localised in the holes through the walls of which the current flows. In other forms of resonator it is not possible to make this distinction, but it is still true that the electric field is concentrated mainly in the region of the open end, whereas the current is highest at the closed end, the magnetic field therefore being strongest in the latter area. (The fact that the capacitance and inductance cannot therefore be clearly distinguished, naturally does not imply that the natural frequency and impedance of the cavity resonator cannot formally be defined by an "effective" capacitance and inductance.)

The number of cavities in a modern magnetron is fairly large, being usually eight or more. In order to deliver the required high current strength, the cathode is given a large diameter. The diameter of the anode is then also large, but the spacing of the cavities, which is roughly equal to (E/B) times the half-cycle, cannot be increased beyond a certain limit. In any case B must be greater than B_{cr} and accordingly cannot be reduced *ad libitum*. E must not be too high either, or we run into impractically high voltage values (50-100 kV). A fairly large number of cavities is therefore unavoidable.

As the cavities are mutually coupled via the end chambers and the central hole, the whole system has a series of natural frequencies, numbering $n+1$ in a magnetron with $2n$ identical cavities.

A system of two coupled circuits is able to oscillate in two different ways; the oscillations may be either in phase or in antiphase similarly, the modes of oscillation in a system of more than two coupled cavity resonators are distinguished by their mutual phase displacements. The phase displacement φ between two successive cavities is always constant; if there are $2n$ cavities, we find $2n$ times that phase difference over the cavities in turn, round to the starting point again, and this value must be a whole multiple of 2π:

$$2n\,\varphi = k\cdot 2\pi\,, \quad \text{or} \quad \varphi = k\frac{\pi}{n}\,,$$

where k is any whole number. We then speak of the mode number k. For every value of k there is a natural frequency; for $k = 0, 1, 2 \ldots n$, there are $n+1$ different natural frequencies. When $k = n$, $\varphi = \pi$ and this is usually referred to as the π-mode. Two adjacent cavities are then in antiphase. When $k > n$, other modes of oscillation occur, but the oscillation $n+m$ has the same natural frequency as $n-m$. Such modes of oscillation are known as degenerate. The occurrence of degenerate modes is due to the circular symmetry of the anode block; the two field configurations are at a given moment images of each other with respect to a given plane through the axis. Under conditions of oscillation which are degenerate, that oscillation may easily be excited that is not coupled with the electronic motion, and for this reason magnetrons are designed to oscillate at the (non-degenerate) π-mode.

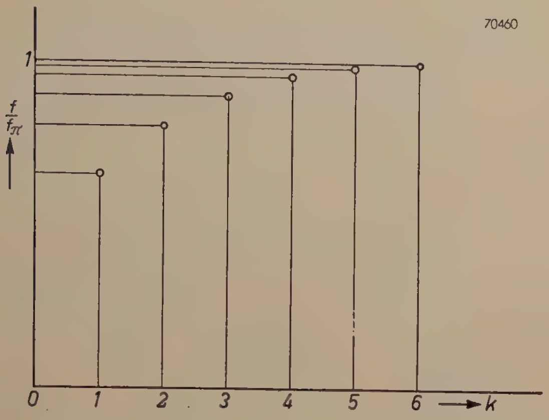


Fig. 15. Natural frequencies of a magnetron with twelve identical cavity resonators. The number k of the mode of oscillation is plotted horizontally and the frequency f vertically, the unit being the frequency f_π of the π-mode ($k = 6$). In the vicinity of the π-mode the frequencies are close together.

Fig. 15 gives an idea of the natural frequencies of a magnetron with twelve cavity resonators; the frequency unit is the frequency of the π-mode. It will be seen that the natural frequencies are very closely spaced in the region of the π-mode, and the risk that the required oscillation will excite the system in such a way as to produce other modes of oscillation is very great. This is detrimental to both the energy transfer and the efficiency, and efforts have therefore been made to find means of increasing the difference between the natural frequencies of the π-mode and neighbouring modes of oscillation.

Improvements in cavity resonator systems

One method of effecting such improvement consists in fitting "straps" between those anode segments which carry the same H.F. voltage, for the π-mode; in other words, all the even segments are strapped with copper wires or rings and likewise the uneven segments. Two methods of strapping are illustrated in *fig. 16*.

From the aspect of the π-mode, these straps increase the capacitance of the segments with respect to each other. Only the charging current of the capacitance flows in the segments, since the extremities are at the same potential. Let the combined effective capacitance of all the cavities be denoted by C_t and that of the straps by C_s; the wavelength is now increased in proportion with $\sqrt{1+(C_s/C_t)}$, which immediately follows from the well-known expression for the resonant frequency f_0 of an L-C circuit, viz:

$$\omega^2 LC = 1,$$

where $\omega = 2\pi f_0 = 2\pi c/\lambda$ (c = velocity of propagation, λ = wavelength).

For other modes of oscillation, the voltage between adjacent segments is lower and the effective capacitance C_s therefore smaller. (The same applies when extra capacitance is added to a transmission line; the effect of this capacitance is greatest at the voltage antinode and decreases towards the node.) Furthermore, current now also flows in the straps, seeing that the extremities are no longer at the same potential. Effective self-inductance then occurs in parallel with the original inductance and thus reduces the overall inductance. Since both capacitance and inductance are lower than in the case of the π-mode, the wavelength is also smaller.

In this way it is possible to make a fairly clear distinction between the natural frequencies of the π-mode and those of other modes of oscillation, amounting to 5% or more, and it is accordingly possible to prevent oscillation in undesirable modes.

Another method of modifying the spectrum of natural frequencies consists in making the cavities of unequal sizes, the most usual

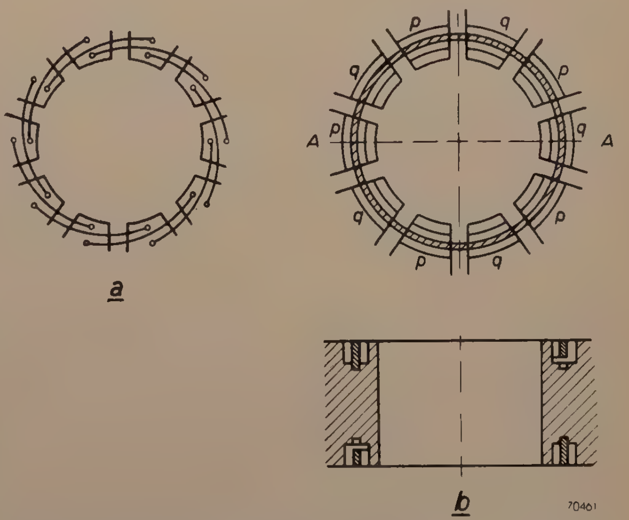


Fig. 16. *a*) "Echelon strapping". Each anode segment is connected to two others by means of a wire.
b) "Ring strapping". The cross section AA is also shown. The ring misses one segment, its hight being reduced at that point, and is attached to the next where it is higher. The ring above the anode block connects the segments p, and that below it the segments q.

arrangement being the "rising sun" system depicted in *fig. 17*, which has ten large and ten small cavities.

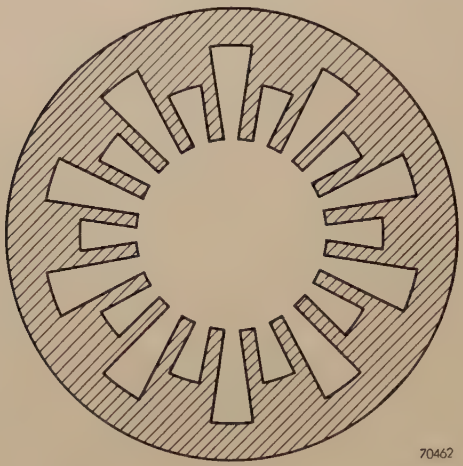


Fig. 17. "Rising sun" anode system with twenty cavities, alternately large and small.

Here, the wavelength of the π-mode lies between two groups, one of which is roughly represented by the natural frequencies of the large cavities only and the other those of the small ones. *Fig. 18* illustrates this in the case of a magnetron having eighteen segments.

It is very difficult — or even impossible — to make strapped magnetrons for wavelengths of less than 1 cm, because the straps would then have to be less than 0.1 mm in thickness. For these millimetric waves, "rising sun" systems are normally used.

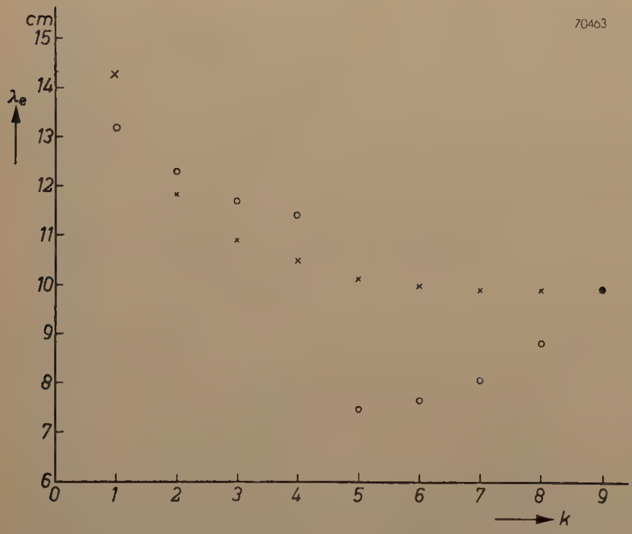


Fig. 18. Natural frequency λ_e as a function of the number k of the mode of oscillation, for two different anode systems consisting of eighteen cavities, which have the same π-mode wavelength. The crosses refer to an unstrapped magnetron and the circles to a "rising sun" system. (From the previously mentioned article in Bell Syst. Techn. J. 25, 229, 1946.)

Output systems

The H.F. power is taken from one of the cavity resonators usually in one of the two following ways.

The first of these employs a loop coupled to one of the cavity resonators and connected to the inner conductor of a coaxial line. Between the inner and outer conductors there is a glass seal which not only provides the output terminal, but also hermetically seals the tube. Often, the coaxial portion is coupled into a waveguide; sometimes, again, the end of the waveguide is soldered direct to the magnetron in the manner shown in *fig. 19*.

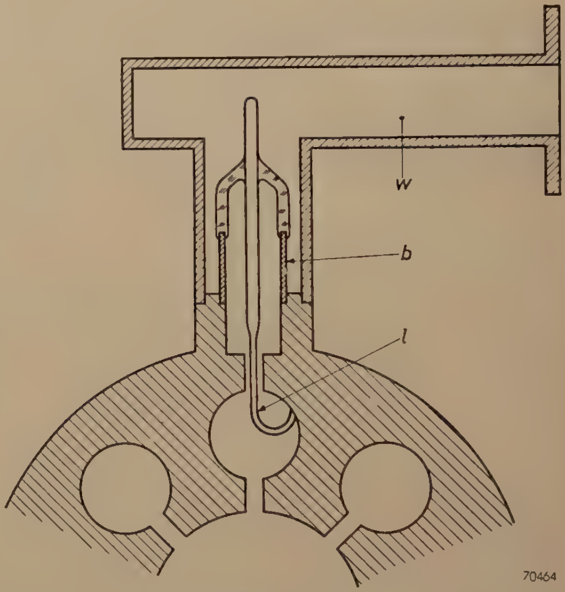


Fig. 19. Coaxial output system. The loop *l* is connected to an inner conductor which is attached to the outer conductor *b* by glass. The waveguide *w* is soldered to the magnetron.

According to the second method a slot is provided in the rear wall of one of the cavities and the power is taken from here direct to a waveguide, which, forming an integral part of the magnetron, also contains a "window" which functions as vacuum seal. This method is illustrated in *fig. 20* and it is employed in many instances, e.g. when the power is so high that there would be some risk of flashover to the probe in the waveguide, owing to the high field strength. Or it may be used when the wavelength is so short that it would be difficult to incorporate a loop in the resonators.

The coupling between resonator and waveguide should preferably be such that the greatest possible amount of power is delivered. It should not be too tight, however, for the frequency shift due to small variations in the impedance of the aerial system must remain sufficiently small. Many radar systems employ a rotating aerial and the rotation always results in some fluctuation in the impedance.

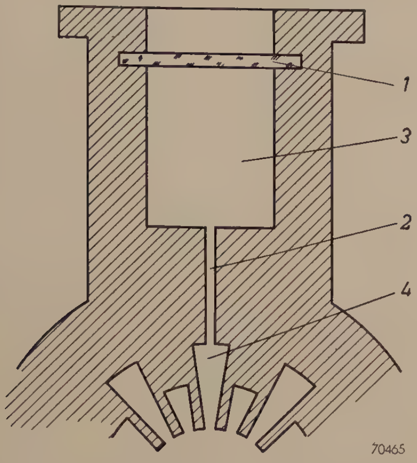


Fig. 20. Cross section of slot coupling (diagrammatic). *1* = vacuum-tight window, *2* = slot, *3* = waveguide, *4* = one of the cavity resonators.

A compromise between the condition for effective transfer of power (tight coupling) on the one hand and that which promotes stability (loose coupling) on the other, is established by suitably dimensioning the coupling gap. In principle it would be possible to provide some matching arrangement outside the magnetron, somewhere in the waveguide, but the great objection to this is that it would produce standing waves in the waveguide, between the magnetron and the matching arrangement; such would tend locally to increase the field strength, which in turn might lead to flashovers. Moreover, it is often required that the magnetron can be replaced quickly, without the necessity of having to make readjustments.

The characteristics of an output system are shown in a so-called Rieke diagram in which the frequency and output power of the magnetron are plotted as functions of the load. Every point in the diagram corresponds to a certain load impedance "seen" from a fixed point in the output line. A rectangular diagram could be employed for this purpose, with the real and imaginary components as co-ordinates, but it is preferable for practical reasons to use the reflection factor and the phase of the reflected wave as variables with polar co-ordinates[10]). The reflection factor and phase can, in fact, be measured directly and, moreover, care can be taken in the design of the H.F. lines and aerials that the reflection factor does not exceed a certain value. The impedance is then kept within a given circle in the Rieke diagram, the importance of which may be explained as follows.

In *fig. 21* the reflection factor is plotted as radius vector and the phase as an angle. Measurements of power and frequency are taken at numerous different values of the impedance (i.e. different sizes of radius vector and angle), and the Rieke diagram is constructed by plotting the curves through the points of constant power or frequency.

It will be seen from *fig. 21* that the power is high in a certain zone, but that the variation in frequency with the impedance is also large. These conditions easily lead to instability. By varying the size of the coupling gap, the zone in question can be displaced towards the centre, or away from it. The amount of frequency shift is usually indicated by the pulling figure, which is in effect the total frequency variation occurring in the magnetron when the impedance value is moved round the circle representing a reflection factor of 0.2 in the Rieke diagram. Both the pulling figure and the output power are reduced when the width of the coupling gap is decreased. The particular compromise adopted depends on the application.

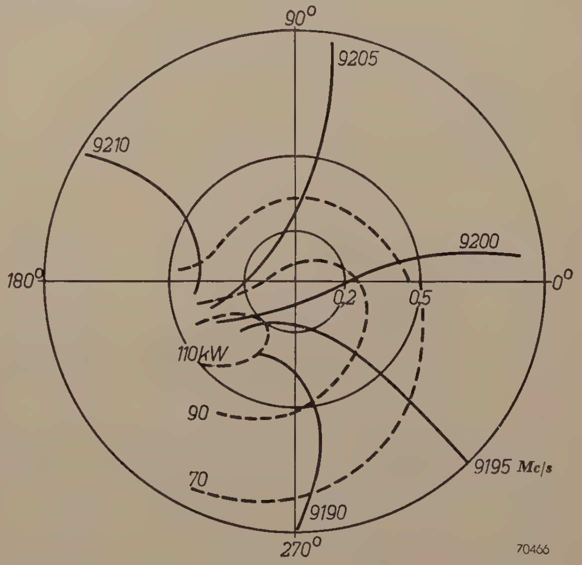


Fig. 21. Rieke diagram of a magnetron for a wavelength of 3 cm. The impedance of the load is plotted in a circular diagram The reflection factor is represented by the radius vector; the phase of the reflected wave at any point is given by the angle between the radius vector and a fixed direction. In this diagram, lines of constant frequency (in Mc/s) and constant power (in kW) are drawn. The overall frequency variation occurring within the circle for a reflection factor of 0.2 is known as the "pulling figure" (in this case approximately 12 Mc/s).

---

[10]) P. H. Smith, Transmission line calculator, Electronics **12**, 29-31, 1939.

Constructional details

The cathode

The cathode of a magnetron has to meet some very high standards. In the first place the current density is great; in pulsed magnetrons cathode loads of more than 30 A/cm^2 are common. Further, at such current densities, in conjunction with anode voltages of 30 kV or more, the cathode must not arc, and it must be possible to dissipate the heat generated by electronic bombardment in the wrong phase (see above). These requirements become so much the more rigid according as the output power is increased and the wavelength reduced.

Oxide-coated cathodes are found to withstand loads of 10 A/cm^2 surprisingly well, provided these are applied in the form of pulses of about 1 μsec duration; a cathode life of 1000 hours or more is by no means unusual, but it should be remembered that if the duty cycle is, say, 1 : 1000, a life of 1000 hours means that the cathode has actually emitted only for one hour.

In the case of tubes for still higher cathode loads use is often made of a cathode having metal gauze in the oxide coating, or a metallic powder sintered over it which is coated with another oxide layer. It is actually found that, if this is not done, heavy currents result in such a large potential difference across the cathode that disruptive discharges are likely to occur in the coating [11]). In consequence, the oxide coating will generally sputter; the gauze or metallic powder more or less short-circuits the coating and thus limits the potential differences.

The L-type of cathode mentioned in an earlier article [12]) is particularly suitable for magnetrons. Not only is the specific emission high, but the resistance of the whole system is, moreover negligible. Arcing is very much less in this type of cathode than in others and the cathode is hardly heated by the pulsatory currents, in contrast with oxide-coated cathodes, in which considerable heat is developed by these currents because of the high resistance of the oxide coating [13]).

In some tubes it is a disadvantage of the L-cathode that it consumes more heater power than an oxide-coated cathode. For the magnetron, however, this is a decided advantage. Once the tube is working, the cathode remains hot when the heater voltage is reduced to zero, owing to the electrons striking it; it is therefore able to dissipate an amount of power equal to the heater power in the form of back-bombardment heat, which is a certain fraction of the input power. Owing to the higher heater power consumed by the L-cathode, the amount of D.C. input power of a magnetron with this type of cathode can also be higher.

In the case of oxide-coated cathodes, good use is often made of the high secondary emission resulting from the returning electrons. Magnetrons with this type of cathode can be operated at a current that is a multiple of the primary emission. Since the coefficient of secondary emission of the L-type of cathode is very much lower, the use of this cathode means that the dependence on the primary emission is greater.

Fig. 22 depicts the L-cathode as employed in a magnetron suitable for a wavelength of 3 cm. The flanges at the ends are provided in all magnetrons to prevent electrons from straying towards the end spaces of the magnetron itself. Currents arising from such electrons cannot produce any H.F. oscillation and thus result only in detrimental heating of the end plates.

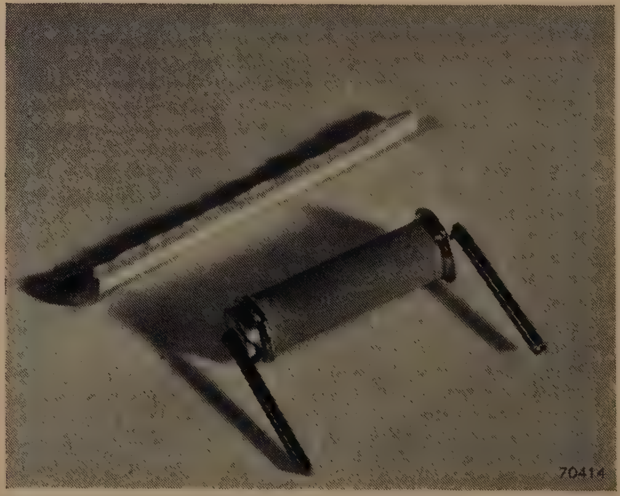


Fig. 22. L-type of cathode for a magnetron for a wavelength of 3 cm.

The anode system

The manufacture of anode systems for magnetrons for wavelengths of 3 cm or less is difficult because of the small dimensions of the system, especially when the magnetron is strapped; the straps, which are not more than a few tenths of a millimetre in thickness, have to be soldered in position very carefully. It will be appreciated that special methods of manufacture have had to be evolved for magnetrons suitable for a wavelength of only 1 cm, the linear dimensions being one third of those of a tube for a wavelength of 3 cm.

---

[11]) R. Loosjes, H. J. Vink and C. G. J. Jansen, Thermionic emitters under pulsed operation, Philips Techn. Rev. **13**, 337-345, 1952 (No. 12).

[12]) H. J. Lemmens, M. J. Jansen and R. Loosjes, Philips Techn. Rev. **11** (341-350, 1950).

[13]) E. A. Coomes, The pulsed properties of oxide cathodes, J. appl. Phys. **17**, 647-654, 1946.

There are several ways of making magnetrons with cavity resonators in the form of sectors. The walls, which divide one cavity from the next, can, for example, be soldered to a copper outer ring with the aid of special jigs or, again, the systems may be produced by electrolytic or casting processes, using a suitable matrix.

We shall, however, describe in more detail another method which is particularly suitable when large number of precisely similar anode systems are to be made, even if these be of extremely small dimensions, the method in question being known as hobbing. The anode shown in *fig. 23* has been made in this way. The "hob" consists of a steel "negative" very accurately ground to size. This "hob", mounted in a press, is forced into a solid block of copper, the height of which is roughly equal to the desired length of the anode, and the copper is thus made to flow into the details of the hob, which is subsequently withdrawn.

The grinding of a tool of this kind is a lengthy process; it can be done automatically and with great precision on a machine which each time rotates the hob to the extent of one slot, after which a high-speed grinding disc grinds out the slot to a certain depth. When the hob has completed one revolution the feed is automatically increased and each of the slots is deepened. The leading end of the hob terminates in an obtuse point and the block of copper to be hobbed is recessed to receive

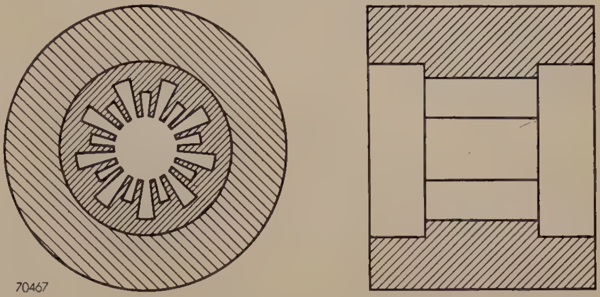


Fig. 23. "Rising sun" anode for a magnetron for a wavelength of 3.2 cm, drawn three-fourths of actual size in cross and longitudinal sections.

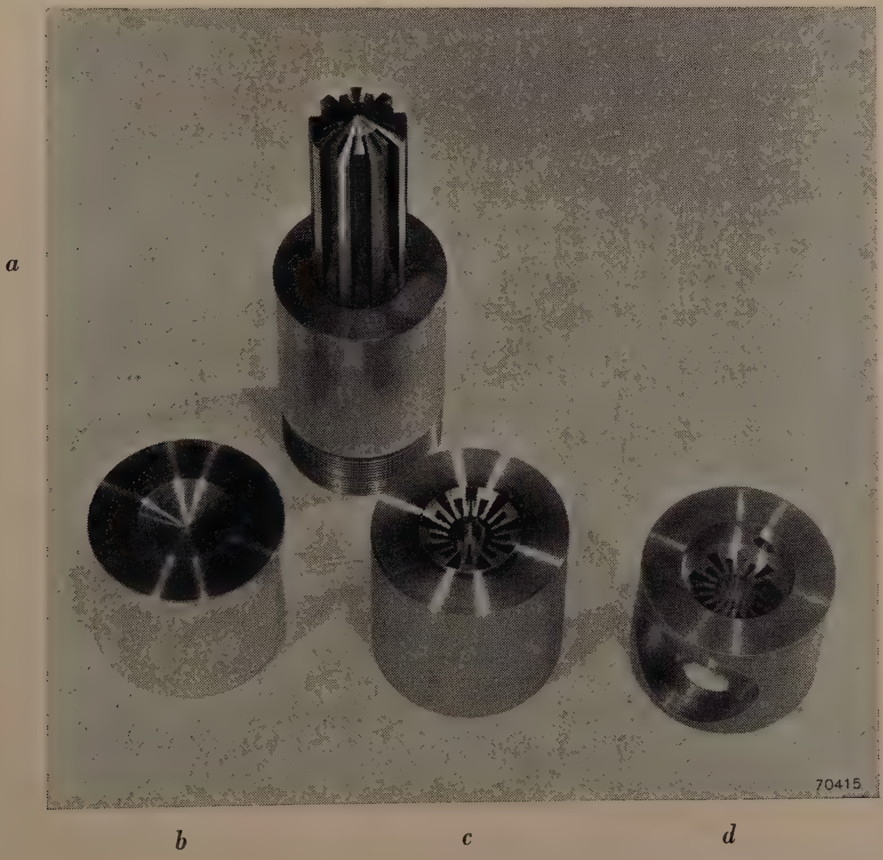


Fig. 24. *a*) Hob. *b*) Copper block to be hobbed. *c*) After hobbing. *d*) After turning. The waveguide is connected to the aperature in the side in (*d*).

this. In *fig. 24* the hob is seen at the rear and, in front of it, from left to right, blocks of copper before and after hobbing, and the same block after having been turned down to the required dimensions for the magnetron. In order to protect the cavity walls from damage during the final machining, the hobbed anode is filled with a thermo-plastic substance known as "Lucite", the essential feature of which is that it sets without the inclusion of any air-bubbles. After the turning operation the "Lucite" is removed by means of a solvent. Some hundreds of anode systems can be made with a single tool of this kind. Incidentally, waveguides and coupling slots can also be produced comparatively easily by the same process.

In conclusion we reproduce in *fig. 25* two photographs of a hobbed magnetron with L-type of cathode, for a wavelength of 3 cm. The performance diagram in fig. 12 and the Rieke diagram fig. 21, relate to this magnetron.

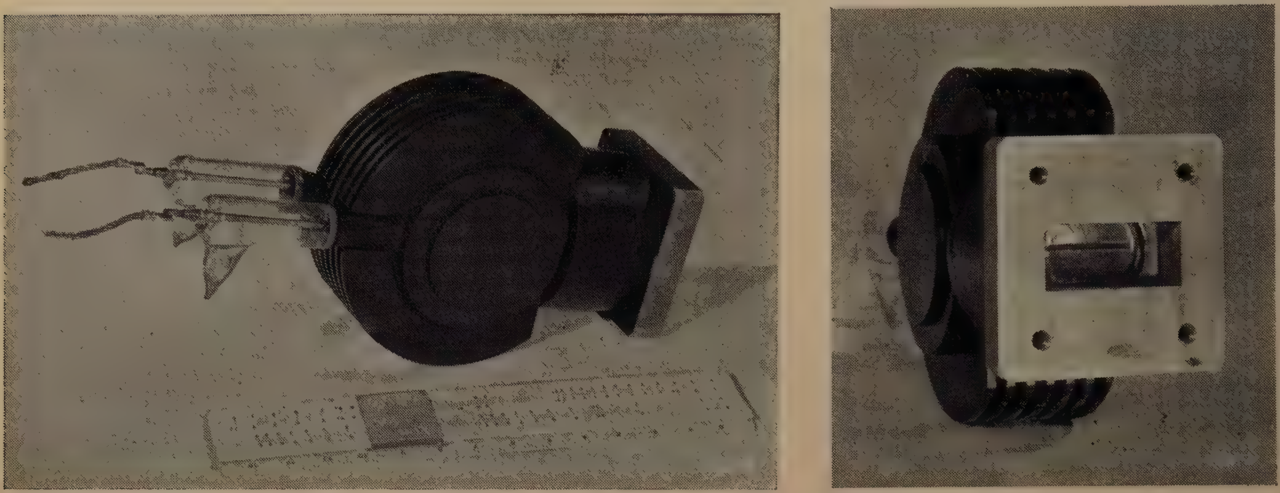


Fig. 25. Magnetron for a wavelength of 3.2 cm, to deliver 1000 kW peak. The right-hand illustration shows the waveguide output connection; the coupling slot is seen through the window. The cathode of this magnetron is depicted in fig. 22 and the anode in fig. 24*d*.

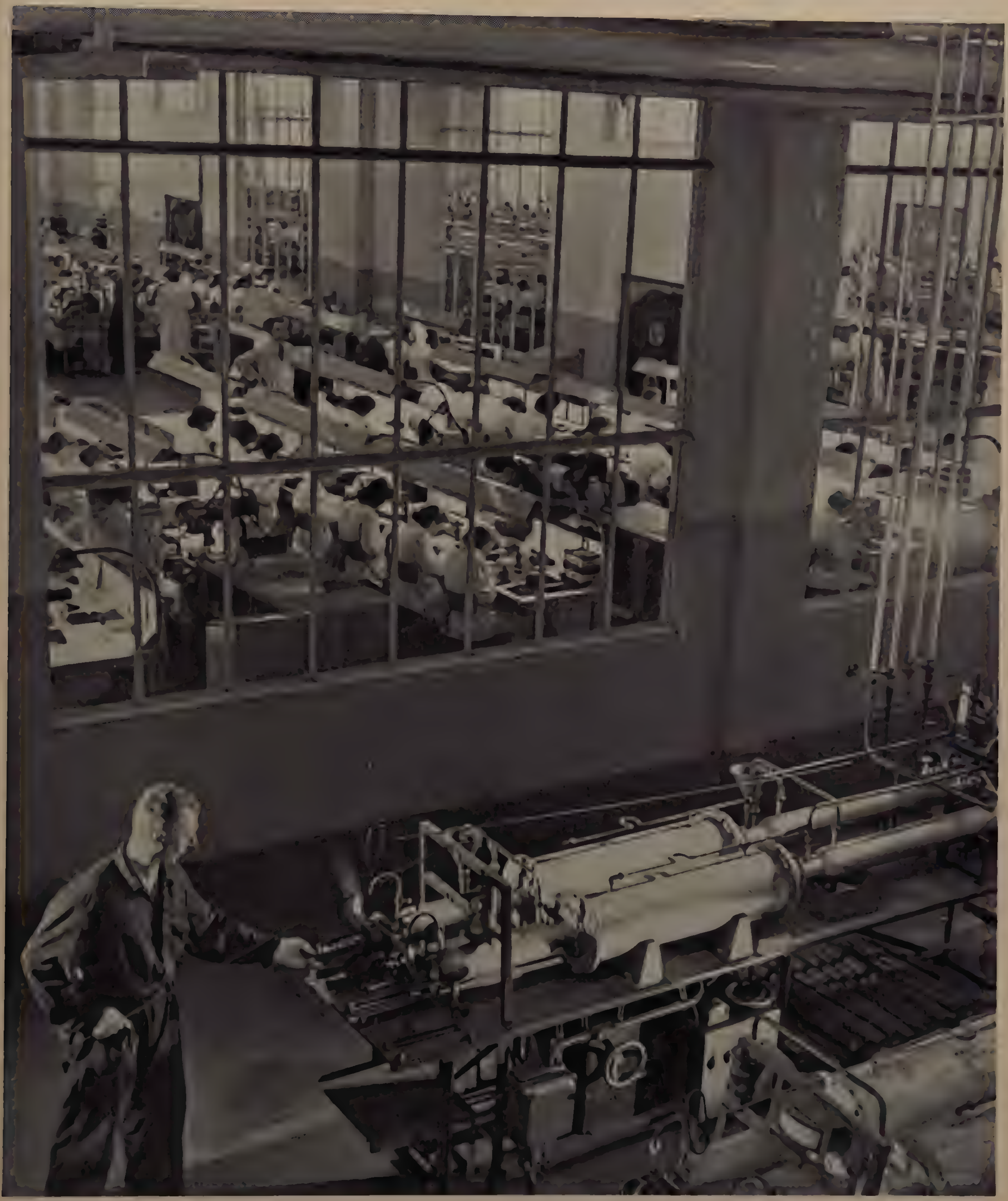
Summary. This introductory article on magnetrons explains the motion of electrons when influenced by mutually perpendicular electric and magnetic fields. A quantitative explanation of the occurrence of oscillations is given, based on the considerations formulated by Posthumus. A brief account is also given of the useful effects of phase focusing, or formation of electron bunches, and of Hartree's method of iteration as applied to the computation of electronic paths. Different relationships between the direct anode voltage V and the flux density B set a limit in the V-B plane to the zone in which oscillation is possible. Mention is made of the performance chart.

The resonators employed with magnetrons now usually take the form of cavity resonators incorporated in the anode. Details of the spectrum of resonant frequencies resulting from the coupling between the cavities are discussed, as well as means of increasing the difference between neighbouring resonant frequencies in order to avoid possible unwanted modes of oscillation (strapping, or the use of the "rising sun" system of alternate large and small cavities).

Output systems for magnetrons generally take the form of a coupling-loop or slot in one of the cavities. The characteristics of such systems (inter alia the pulling figure) are expressed in a Rieke diagram, an example of which is given.

Details regarding the design of magnetrons include the stringent requirements to which the cathode must conform (these being adequately met by the L-type of cathode), and the method of manufacture of the anode block with cavity resonators by the hobbing process.

PART OF THE PHILIPS VALVE FACTORY AT SITTARD (HOLLAND)



Photograph Walter Nürnberg

STRAIN AGEING IN IRON AND STEEL

by J. D. FAST.

620.193.918.4:669.11:669.14

There is no need to stress the importance of fundamental research into the influence of impurities in technical metals such as steel. Impurities are responsible, i.a., for the ageing phenomena arising in iron and steel after cold working (rolling, drawing). Experiments carried out with alloys of well-defined composition show once more the important part played particularly by nitrogen as an impurity.

Ageing

Steel, and particularly so-called mild steel, shows ageing phenomena following upon mechanical deformation (rolling, drawing). Such phenomena manifest themselves mainly in a gradual increase in the hardness of the material after deformation. This gradual change in the hardness of iron and steel takes place also after rapid cooling (quenching) from a high temperature. In the former case one speaks of strain ageing, in the latter case of quench ageing. An article on quench ageing appeared in this journal a short time ago [1]), when it was explained that this phenomenon is due to the precipitation of particles of iron carbide or iron nitride in the metal, such arising from the presence in the metal of quantities of carbon and nitrogen in the dissolved state prior to the quenching; pure iron does not show any quench ageing. The reader is reminded here of the remarkable phenomenon that a small quantity of manganese appeared to delay quench ageing of nitrogen-containing iron considerably.

Strain ageing, too, appears to be caused by particular constituents in the metal. In *fig. 1* curves are given showing the change in hardness taking place in course of time after quenching and after deformation of mild steel (according to experiments by Davenport and Bain [2])). The broken lines represent the change in hardness after quenching from 720 °C to room temperature and subsequent storage at temperatures of 40 °C and 100 °C respectively. It appears that the hardness first increases and then decreases. The aforementioned precipitation of iron carbide and iron nitride in the form of small particles causes the increase in hardness, the subsequent decrease being due to coagulation (coarsening) of the precipitated particles. As the curves show, the higher the temperature the more rapidly both these changes take place.

The fully-drawn curves show the change in hardness after the material has been subjected to a deformation of 10% at room temperature (cold working). In this case, too, there is at first an increase in hardness during storage, the increase being more rapid the higher the temperature, but there is virtually no decrease in hardness as time goes on. Thus in this respect there is a decided difference between strain ageing and quench ageing.

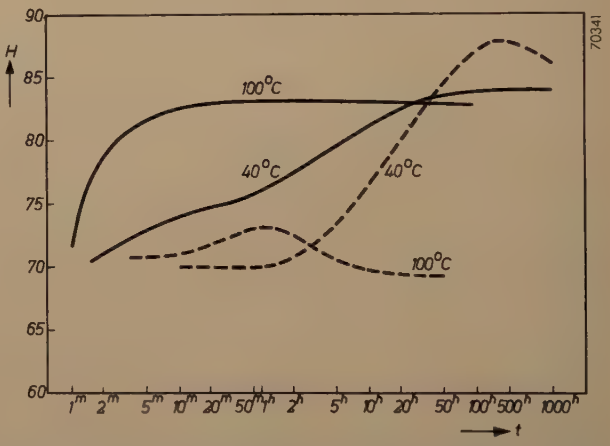


Fig. 1. Quench ageing and strain ageing of mild steel, according to Davenport and Bain [2]). Vickers hardness H as a function of the time t at 40 °C and 100 °C, after quenching from 720 °C (broken-line curves) and after 10% cold deformation (fully drawn curves). The hardness of the quenched material first increases (precipitation) and then decreases (coagulation). The hardness of the deformed material increases, but at the temperatures given does not subsequently decrease. Note the difference in maximum hardness at the two temperature in the case of the quenched material. For the deformed material the final value of the hardness is practically independent of the ageing temperature.

Another effect of strain ageing is to be seen in the stress-strain curve of technical iron. This curve represents the relation between the load (stress) to which the rod-shaped material is subjected and the deformation (strain) resulting therefrom. In *fig. 2a* a stress-strain curve is given for mild steel.

---

[1]) J. D. Fast, Ageing phenomena in iron and steel after rapid cooling, Philips Techn. Rev. **13**, 165-171, 1951 (No. 6). This article will further be referred to as I.

[2]) E. S. Davenport and C. E. Bain, Trans. Amer. Soc. Met. **23**, 1047, 1935.

Under small loads there is an almost linear relation between the elongation (deformation) and the load, but when a load is reached corresponding to the point *A* plastic deformation of the material begins to take place, the deformation then increasing very suddenly and rather irregularly, while at the same time the load decreases somewhat as a result of the inertia of the tensile testing machine. Finally, beyond *C*, any further plastic straining of the material requires again an increasing load. The stresses corresponding to the points *A* and *B* are called respectively the upper and lower yield points of the steel.

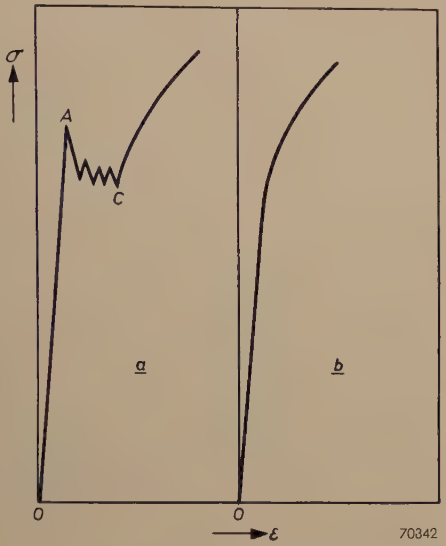


Fig. 2. Stress-strain curves of mild steel. Stress σ plotted as a function of the deformation ε. *a*) The steel shows an upper yield point at *A*. Material first strained as far as *C*, shows under renewed loading a smooth stress-strain curve (*b*).

When the load is removed as soon as the point *C* in fig. 2*a* is reached and the stress-strain curve is recorded anew, the smooth curve of fig. 2*b* is obtained. The yield points have disappeared. Strain ageing then manifests itself in a gradual return of the sharp yield points when, following upon the first treatment (deformation up to the point *C*), the material is left to itself, and the higher the temperature at which the material is kept the more quickly this return takes place.

Neither of these two effects of strain ageing — gradual hardening after deformation and gradual return of the two yield points — occurs in pure iron (which as a matter of fact shows no yield points at all). It is therefore to be concluded that strain ageing as well as quench ageing is related to impurities which are always present in technical iron. But whereas the cause of quench ageing has been found to lie in the formation of precipitated particles, according to Cottrell [3]) strain ageing is to be ascribed to the behaviour of the carbon and nitrogen atoms remaining in the dissolved state.

To check this assumption, hardness tests have been carried out with pure iron and with iron to which known quantities of one or more impurities had been added by a method described earlier in this journal [4]); as impurities, carbon (0.04%) and nitrogen (0.02%) were used, with and without manganese (0.5%). Before proceeding to discuss the experiments it is well to consider briefly the explanation suggested by Cottrell *et al* for strain ageing, since the results of our experiments seem to make it necessary to reconsider some details of that theory.

Plastic deformation of a metal

The plastic deformation of a metal consists in parts of crystals shearing over each other along certain crystallographic planes (slip planes) and in certain crystallographic directions (slip directions) under the influence of external forces. *Fig. 3* shows, diagramatically, how this might happen. The broken line represents a slip plane perpendicular to the plane of the drawing, the slip direction being along the broken line. Under the influence of the shear stress indicated by arrows, shearing takes place in the manner represented in figs 3*a* to 3*c*. In fig. 3*c* the shear has progressed over the distance of one atom, the atoms meanwhile having passed through positions (fig. 3*b*) in which they have a greater energy than before or after the slip. This energy can be calculated from the crystal characteristics

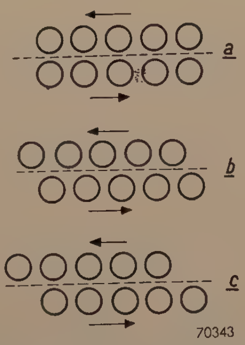


Fig. 3. How two layers of atoms in an imaginary simple cubic metal lattice glide continuously one over the other. To reach the new position of equilibrium *c* from the position *a* the atoms have to pass through the configuration *b*, which has a much greater energy.

---

[3]) A. H. Cottrell, Report of a conference on strength of solids, held at Bristol, 1948, pp. 30-38; Progress in Metal Physics **1**, 77-126, 1949.

[4]) J. D. Fast, Apparatus for preparation of metals with an exactly known content of impurities, Philips Techn. Rev, **11**, 241-244, 1949.

and from that it can be deduced what shear stress would be needed to bring about a displacement of all atoms simultaneously, as represented in fig. 3. Such a calculation gives stresses more than a thousand times greater than those found experimentally, so that apparently the atoms are not all displaced simultaneously. The shear might be imagined as taking place in steps, the atoms being displaced in the slip direction one by one, in which case a much smaller shear stress is required. Represented in the most primitive form, the situation then changes from that in fig. 3 to that in *fig. 4*.

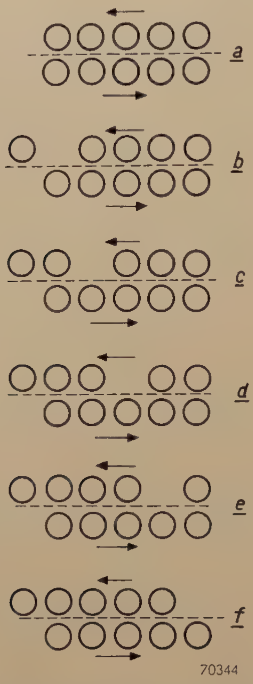


Fig. 4. Simplified representation of the discontinuous gliding of one layer of atoms over another in a simple cubic lattice. The arrows in this and the previous diagram denote the direction of the shear stress. The broken line represents the slip plane.

The representation in fig. 4 is rather too simple, since only those atoms are considered which lie immediately above and below the slip plane. When taking also the other atoms into account we get the picture of *fig. 5*, where the shear has progressed half-way through the crystal (thus the situation as given in fig. 4*c* or fig. 4*d*). The deviation from the ideal structure as shown in the middle of fig. 5 (the outlined part) is called a **dislocation**. It is to be borne in mind that here only a cross section through the crystal is shown and that the dislocation is assumed to extend perpendicularly to the plane of drawing on either side. The dislocation as it exists in the middle of the diagram, while proceeding from left to right, can also be imagined as being formed locally owing to an extra, vertical, layer of atoms perpendicular to the paper (indicated by the short arrow) being forced into the lower half of the crystal. As a result the atoms lie too close together in the lower half of the dislocation and too far apart in the upper half, the lower half being under a compressive stress while the upper half is under a tensional stress. In addition, shear stresses arise round the dislocation.

To the left of the dislocation a slip has already taken place over the distance of one atom, while to the right of it there is as yet no change.

In this picture plastic deformation is nothing but the movement of dislocations along the slip planes in certain crystallographic directions. Under the influence of a shear stress this movement takes place at a high speed, being promoted by the thermal motion of the atoms. When a dislocation has been displaced from one side of the crystal to the other, a shear has taken place over the distance of one atom. Viewed from above a slip plane, the shearing process takes place as diagramatically represented in *fig. 6*. The (hatched) dislocation must either extend over the entire width of the slip plane from one side of the crystal to the other or else form a closed loop.

It is commonly assumed that there are always dislocations already present in the metal crystals. How it is to be imagined that in the formation of the crystals from a melt or from a gas phase these dislocations become incorporated in the solid substance, and how they can be formed at the walls of a crystal during the process of deformation, are

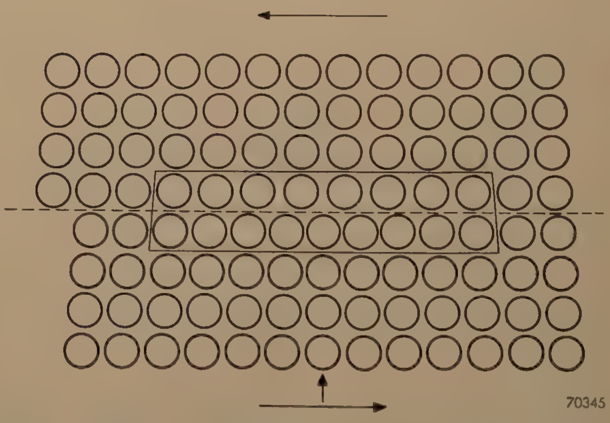


Fig. 5. Schematic representation (in one atomic plane) of a dislocation in a simple cubic metal lattice. The horizontal arrows denote the direction of the external shear stress, the broken line representing the slip plane. In the outlined area the influence of the dislocation is most manifest. The dislocation can also be imagined as being caused by the forced entry of an extra row of atoms in the lower half of the crystal (short arrow). The dislocation is a practically linear lattice disturbance perpendicular to the plane of drawing; here only the cross section with one atomic plane is to be seen.

questions which we shall not enter into here. Anyhow it must be assumed there is a great increase in the number of dislocations during the plastic deformation, and this accounts for the hardening of the metal as a result of the deformation. Each dislocation being surrounded by a stress field, the dislocations exercise attractive or repelling forces upon each other, and as they increase in density the movement of each individual dislocation is rendered more difficult by these forces. Thus the metal becomes less malleable and its hardness increases. This hardening is clearly expressed in the stress-strain curves of fig. 2.

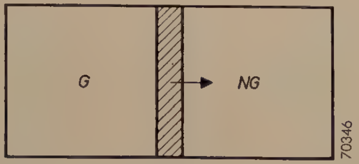


Fig. 6. Displacement of a dislocation from left to right along a slip plane, "seen" in a direction perpendicular to that of the three perceding diagrams, viz. in the direction of the short arrow in fig. 5. *G* is the part already sheared ("the part where the dislocation has already glided over"), *NG* is the part not yet sheared.

According to the dislocation theory the increase in the hardness of a metal after quenching is due to the movement of the dislocations being hampered by the precipitated particles, this probably being caused mainly by the elastic deformation of the lattice surrounding the grains of the precipitate.

Theoretical considerations on strain ageing

According to Cottrell *et al*, strain ageing results from the interaction between the dislocations and the carbon and nitrogen atoms interstitially dissolved in the iron. As already explained in article I, each dissolved carbon or nitrogen atom increases the volume of the elementary cell of the iron lattice in which it is contained. Now in the foregoing we have seen that on one side of a dislocation the iron lattice is stretched and on the other side compressed. Consequently the interstitial places are no longer of equal energy: the dissolved atoms escape from the compressed areas and tend to seek the expanded areas. In the state of equilibrium at room temperature only few atoms are available for escaping from the compressed regions, since the solubility of carbon or nitrogen in iron at room temperature is very small indeed, only a few times 10^{-7} and 10^{-5} wt% respectively. Therefore the interaction between the dislocations and the dissolved atoms will mainly be such that in course of time in the expanded part of each dislocation a concentration of C or N atoms much greater than that in the undisturbed lattice will be established. This results in a decrease of the tensile and compressive stresses surrounding the dislocation. The shear stresses, too, may be reduced by displacements of the dissolved atoms, since each C or N atom causes an asymmetric deformation of the iron lattice (see I). This may all be briefly expressed by saying that the elastic energy of the dislocation (or rather, of the stress field around the dislocation) is reduced.

In the picture of fig. 5 the cloud of C or N atoms formed in and around a dislocation might consist mainly of one central row of atoms, i.e. at most one C or N atom per atomic plane parallel to the plane of drawing. Around that "cloud" the concentration decreases exponentially. Such strings of C or N atoms may perhaps be regarded as an intermediate stage between the dissolved and precipitated states of carbon (nitrogen). It might be assumed that these one-dimensional pre-precipitates act as nuclei in the precipitation proper.

As a consequence of its reduced elastic energy the dislocation becomes, as it were, anchored, and thus the hardness of the material is increased. In this way strain ageing may be considered as a process of diffusion of the interstitially dissolved carbon or nitrogen atoms towards the dislocations.

In a technical steel that has not recently been deformed a large portion of the dislocations will always be anchored. The displacement of such a dislocation will increase its elastic energy, unless it is able to take along with it its cloud of interstitial atoms. When at a relatively low temperature a comparatively rapidly increasing load is applied to the material, the clouds will not have time to diffuse and as the load increases the dislocations will finally be "torn away" from their clouds, after which a smaller stress is required for the further movement of the dislocations. To put it in other words, a greater stress is required to initiate the plastic flow than to keep the material flowing. Thus the theory gives a natural explanation for the occurrence of the two yield points in technical iron (see fig. 2*a*).

If after the first overall yielding the iron is not further deformed and the stress-strain curve is again recorded after not too long a time, the dissolved atoms will not yet have had sufficient opportunity to diffuse towards the dislocations again, so that the latter will then still be able to move easily. This explains why no yield points are found in the new stress-strain curve. If, however, one waits so long, after the recording of the first curve, that the interstitial atoms have had sufficient time to diffuse towards the dislocations and thus to anchor them, then it is to be expected that an additional stress will again be needed to start plastic flow, in other words, that the two yield points will have returned. As already stated, this is exactly what is observed. The higher the temperature, the greater will be the rate of diffusion and the sooner will be the return of the yield points. This, too, is confirmed experimentally, as we have already seen.

The theory is quantitatively confirmed by the fact that the temperature coefficient for the rate of this strain ageing in carbon-containing iron appears to be about equal to the temperature coefficient for the diffusion constant of carbon in iron (corresponding to an activation energy of about 80 000 joules per mole, see I).

Strain-ageing experiments with specially prepared alloys

It was considered well worth while to investigate whether the above-mentioned views held by Cottrell and others, which are mainly based upon experiments with technical iron, also hold in the case of pure iron to which known quantities of one or more impurities have been added. In the course of the experiments carried out with these materials [5]) it appeared that carbon and nitrogen do not have exactly the same effects. Even after forty hours' heating at 50 °C carbon did not give any strain ageing, whereas nitrogen gave the maximum effect already after two hours' heating at that temperature. The differences are shown graphically in *figs* 7 and 8. With carbon as the only impurity the iron has to be heated to 200 °C (heating time two hours) before the maximum ageing effect is obtained.

The procedure for these experiments was analogous to that followed for the quench-ageing experiments described in article I. First the hardness of the material was measured immediately after the deformation and then, for the ageing phenomenon to take place within a reasonable time, the material was kept at 50 °C for two hours, after that at 100 °C for two hours, at 200 °C for two hours and so on up to 700 °C, taking measurements of the hardness after each period of heating. The advantage of this procedure is that the last stage of the ageing, which at low temperatures sometimes takes a very long time, now takes place at higher temperatures, thus requiring less time.

Figs 7 and 8 show that after the heat treatment at 300 °C the hardness of the material decreases again. This would appear to contradict our statement that, as opposed to what takes place in quench ageing, in the case of strain ageing there is no loss in hardness after long storage. This, however, is only a seeming contradiction. At temperatures above 300 °C the processes of recovery and recrystallization come into play and it is this that causes the hardness to be reduced. The descending part of the curves in figs 7 and 8 is therefore due to processes other than those dealt with here.

The experiments described indicate that Cottrell's theory, according to which strain ageing of technical iron is caused by a regrouping around the dislocations of carbon or nitrogen atoms which are already dissolved in the iron, has to be modified slightly.

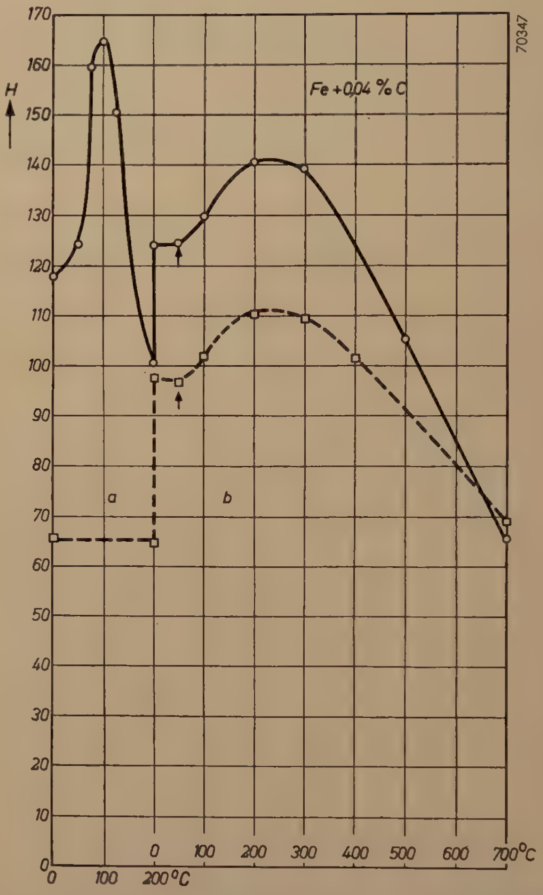


Fig. 7. The Vickers hardness H of iron containing 0.04% carbon as the only impurity, after different thermal and mechanical treatments. The fully-drawn curve gives the values of the hardness after the following treatments: quenching from 720 °C followed by two hours' heating successively at 50, 75, 100, 125 and 200 °C, 10% deformation (elongation), two hours' heating at 50, 100, 200, 300, 500 and 700 °C. The broken curve gives the hardness after deformation alone and heating at the various temperatures for two hours. In neither case is there any perceptible strain ageing (see the small arrows) during the two hours' heating at 50 °C. The part a of this diagram relates to the change in hardness after quenching, part b to the change in hardness after cold deformation.

Cottrell's theory can certainly account for the fact that at low temperature carbon does not immediately cause strain ageing. It is known that carbon is much less soluble in iron than nitrogen: the solubility of nitrogen in iron is a about 2×10^{-5}% at 20 °C, whereas in the case of carbon this value is only reached at 150 °C. Apparently, therefore, below about 100 °C there are not sufficient carbon atoms in solution for the dislocations to be anchored.

The fact, however, that at a somewhat higher temperature, say 100 °C, carbon already gives rise to a relatively rapid strain ageing, indicates that atoms must then be available for anchoring dislocations, whilst the slightly increased solubility at that temperature is certainly not sufficient to account for the effects observed. It is, therefore, to be concluded that the carbon atoms required for anchoring the dislocations are supplied for the greater part by the carbide particles. Apparently these carbon atoms pass into solution and then through diffusion move towards the dislocations. This would imply that from the point of view of energy it is at least just as advantageous for the C atoms to be in a dislocation cloud as in a carbide crystal.

---

[5]) J. D. Fast, Revue de Métallurgie 47, 779-786, 1950.

According to this modified theory the ageing in carbon-containing iron will take place at the same rate as that in nitrogen-containing iron at 50 °C, only at that temperature where the product of the diffusion coefficient and the solubility of carbon is equal to that for nitrogen at 50 °C. It is thereby presumed that the distances between the dislocations and the grains of the precipitate in the carbon-containing iron are about equal to those in the nitrogen-containing iron, and that the transition of a C atom from the precipitate to the solution requires about the same activation energy as the corresponding transition of an N atom.

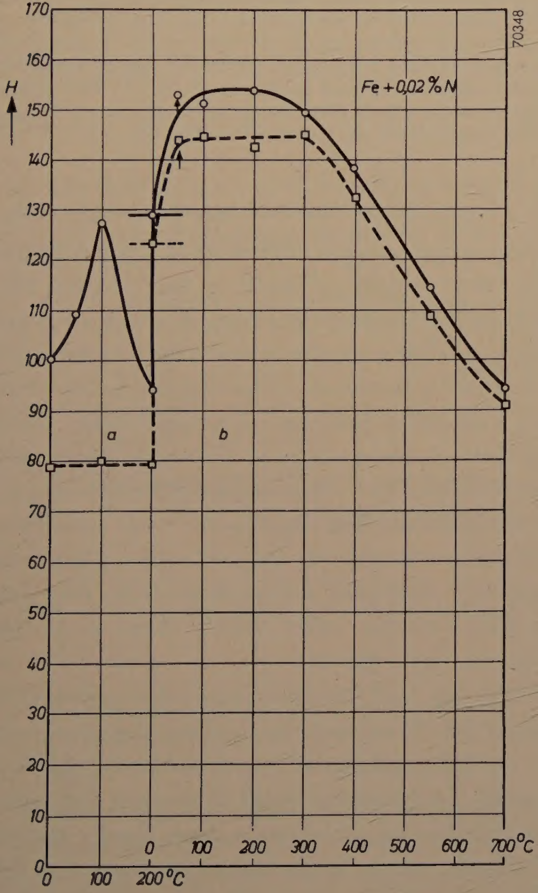


Fig. 8. Vickers hardness H of iron with 0.02% nitrogen as the only impurity, after the thermal and mechanical treatments defined in the subscript of fig. 7. The parts denoted by a and b have the same significance as in fig. 7. Strain ageing takes place very much quicker with nitrogen than with carbon as impurity; the maximum hardness is already reached after two hours' heating at 50 °C (small arrows). The decrease in hardness at higher temperatures in these two diagrams is due to recovery and recrystallization.

According to the most recent figures 6) the solubility q and the diffusion coefficient D of carbon and of nitrogen in alpha iron are given by the formulae:

$$q_C = 3{,}1 \; e^{-10\,000/RT} \text{ wt } \%,$$
$$q_N = 5{,}0 \; e^{-7\,100/RT} \text{ wt } \%,$$
$$D_C = 0{,}02 \; e^{-20\,100/RT} \text{ cm}^2 \text{ sec}^{-1},$$
$$D_N = 0{,}003 \; e^{-18\,200/RT} \text{ cm}^2 \text{ sec}^{-1},$$

where T stands for the absolute temperature and R is the gas constant per mole. According to these formulae the product $D{\cdot}q$ for nitrogen in alpha iron at 50 °C is

$$D_N q_N \, (50\ ^\circ\text{C}) = 1{,}4 \cdot 10^{-19} \text{ cm}^2 \text{ sec}^{-1}.$$

The product $D_C q_C$ reaches this value at a temperature of about 100 °C. Actually, however, the ageing in carbon-containing iron at 100 °C still takes place at a slower rate than that in nitrogen-containing iron at 50 °C (figs 7 and 8). This could be explained by assuming that more nitrogen was dissolved in the iron than corresponded to the state of equilibrium, or, in other words, that we had to do with supersaturated solutions of nitrogen, in contrast to what was the case with the carbon-containing specimens. Under such conditions the nitrogen could anchor the dislocations much sooner than the carbon, which first has to pass into solution from the carbide particles. Another possible explanation is that the transition of a nitrogen atom from the precipitate to the solution takes place more readily than the corresponding transition of a carbon atom.

In the foregoing calculation q_N relates to the solubility of nitrogen in alpha iron which is in equilibrium with the stable nitride phase Fe_4N. In the material that was used for the ageing experiments the nitride was actually in the form of the metastable phase $Fe_{16}N_2$ (cf. I). The solubility of nitrogen in alpha iron which is in equilibrium with the latter phase is greater and according to Dijkstra 6) is given by

$$q'_N = 32 \; e^{-7800/RT}.$$

Applying this expression in the above arguments, the temperature of about 100 °C referred to becomes about 105 °C, so that the result changes but little.

---

6) For the solubility figures see L. J. Dijkstra, Tr. Am. Inst. Min. and Met. Engrs **185**, 252-260, 1949, and for the diffusion coefficients see C. A. Wert, Phys. Rev. **79**, 601-605, 1950 and J. Appl. Phys. **21**, 1196-1197, 1950.

The experiments lead to the conclusion that as an impurity in mild steel nitrogen is to be regarded as being very much more dangerous than carbon. This is also borne out by the fact that by adding aluminium a technical steel can be produced which shows very little strain ageing (e.g. the German Izett steel). Aluminium binds only the nitrogen contained in the steel, and not the carbon [7]).

The effect of manganese upon strain ageing

As described in I, the addition of 0.5% manganese to a nitrogen-containing iron delays its quench ageing considerably. This was accounted for by the fact that the dissolved nitrogen atoms become strongly bound to the manganese atoms in the iron lattice, so that they cannot readily precipitate [8]). But, remarkably enough, the results of our experiments on strain ageing have shown that manganese has no perceptible effect upon the strain ageing of nitrogen-containing iron. From the experiments on quench ageing it would have been expected that manganese prevents the dissolved nitrogen from diffusing to the dislocations and that consequently nitrogen would hardly be able to cause strain ageing in an iron containing manganese.

It is difficult to reconcile this phenomenon with the foregoing considerations. A direction in which the solution might be found lies in a possible interaction between the manganese atoms and the dislocations, such that after the deformation a considerable portion of the manganese atoms is present in the dislocations. It is also possible that our explanation for the effect of manganese upon quench ageing is not quite correct and that the delaying action of manganese in the formation of nitride nuclei is to be accounted for in some other way than by supposing that the nitrogen atoms are unable to release themselves readily from the manganese atoms.

Blue brittleness

Some of our experiments concerned with the problem of blue brittleness are even more difficult to fit into the dislocation picture. By the blue brittleness of technical iron (mild steel) is to be understood the fact that at 200 °C—300 °C this material shows a smaller elongation and reduction of area and a greater tensile strength than at lower or higher temperatures (see *fig. 9*). When the test is carried out at a faster rate this state of relative brittleness is shifted to a higher temperature. An example of such a test is the impact test [9]), where a notched bar of given dimensions is broken with a hammer and the labour required to do that is measured. The impact value of technical iron as a function of temperature shows a minimum at temperatures between about 400 °C and 500 °C. Our experiments (described in the article [9])) show that there is no such minimum at all in the curve for the impact value when carbon is the only impurity in the iron. With nitrogen as impurity, on the other hand, there is indeed a minimum (see *fig. 10*).

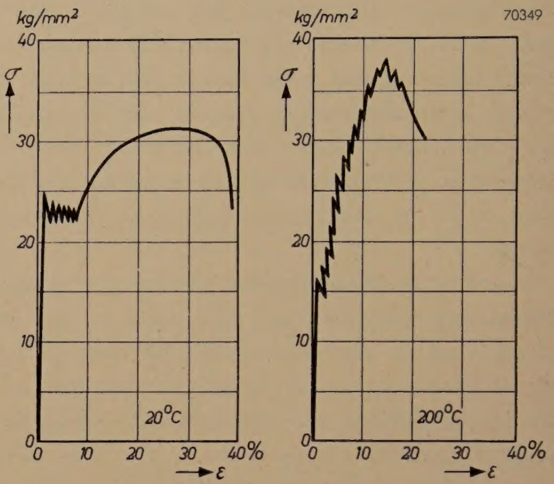


Fig. 9. Stress-strain curves of a non-killed mild steel, taken at 20 °C and at 200 °C respectively. The material shows less elongation and greater tensile strength at 200 °C than at 20°C. The name of blue brittleness given to this phenomenon is related to the heat tinting occurring at 200 °C. The elongation has been measured as the change in length in percents of a test bar originally 5 cm long.

With the aid of the dislocation theory [10]) an attempt has been made to explain the blue brittleness and the minimum in the curve for the impact value in the following way. When the deformation takes place very quickly or very slowly the dislocations experience little hindrance from the clouds surrounding them. When the rate of deformation is very fast the clouds are unable to keep pace with the dislocations and thus no longer have any effect after the latter have broken away from them, whilst in the other case, when the rate of deformation is very slow, the clouds have little effect because they can quite easily keep pace with the movement of the dislocations. There are, however, intermediate rates of deformation where the clouds are still able to move with the dislocations, but only with difficulty. In this range the dislocations are impeded in their movement, and this finds expression in reduced ductility of the material. The higher the temperature, the greater is the diffusion rate of the C and N atoms and the easier the clouds are able to keep pace with the dislocations. As the temperature rises the region of brittleness is therefore shifted to higher rates of deformation, or, what amounts to the same thing, the quicker the deformation takes place the higher is the temperature at which brittleness occurs. At first sight this explanation would seem to be acceptable. It is not clear, however, why nitrogen should cause the material to become brittle and carbon not.

---

[7]) According to investigations by S. T. Epstein, H. J. Cutler and J. W. Frame, J. Metals, June 1950, pp 830-834, strain ageing can also be suppressed by adding vanadium instead of aluminium to the steel. Vanadium combines with carbon as well as with nitrogen, but the object aimed at was reached with much less vanadium than would have been required to bind all the carbon in addition to the nitrogen.

[8]) J. D. Fast and L. J. Dijkstra, Internal friction in iron and steel, Philips Techn. Rev. **13**, 172-179, 1951 (No. 6).

[9]) See J. D. Fast, Investigations into the impact strength of iron and steel, Philips Techn. Rev. **11**, 303-310, 1950.

[10]) See also F. R. N. Nabarro, Report of a conference on strength of solids, held at Bristol 1948, pp 38-45, and G. Masing, Arch. Eisenhüttenwesen **21**, 315-325, 1950.

Although, therefore, the dislocation theory gives many useful indications about the influence of certain impurities in technical iron, in its present form it does not yet explain the details of the behaviour of the impurities.

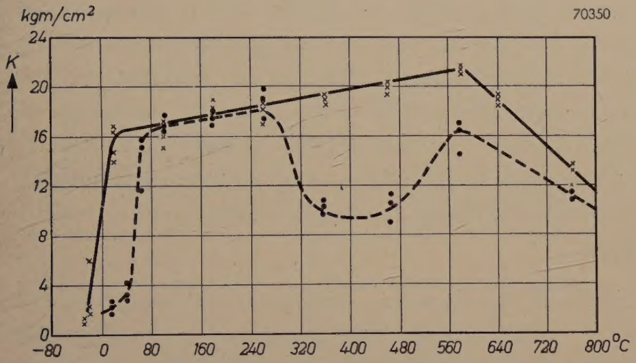


Fig. 10. When carbon is the only impurity in the iron (the fully drawn curve relates to iron with 0.02% C) no minimum occurs in the curve representing the impact value K as a function of temperature. On the other hand, if both carbon and nitrogen are contained in the iron (the broken-line curve relates to iron with 0.06% C and 0.018% N) there is a minimum in the impact value curve between 400 °C and 500 °C.

Summary. By strain ageing of iron and steel is understood a gradual increase in hardness after cold working and the reappearance of the two yield points in the stress-strain curve after its disappearance through pre-treatment. The ageing phenomena are to be ascribed to impurities such as carbon and nitrogen contained in the metal. With the aid of the theory of dislocations attempts can be made to explain these phenomena (Cottrell), but the difficulty is that in the state of equilibrium the amounts of dissolved carbon and nitrogen are too small to cause the effects observed. It is therefore to be concluded that carbon and nitrogen pass into solution from precipitated particles of carbide and nitride respectively and then flow towards the dislocations. Experiments with carefully prepared alloys containing only carbon or nitrogen as impurity show that even after forty hours' heating at 50 °C carbon still does not cause any strain ageing, whereas already after two hours' heating at that temperature nitrogen has the maximum effect. Apparently nitrogen has a much greater effect upon the properties of iron than carbon, as is also apparent from experiments concerning the blue brittleness and the impact value of steel. As far as strain ageing is concerned, manganese does not affect the behaviour of nitrogen, a very remarkable phenomenon, considering that quench ageing is indeed greatly influenced by manganese.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk * can be obtained free of charge upon application on the back cover.

1987: W. de Groot: A graphical method for solving problems of low-level photometry (Appl. sci. Res. B 2, 131-148, 1951, No. 2).

A certain level of subjective brightness may be characterised by the minimum power per unit area and solid angle, E_{m}, with which the corresponding brightness impression can be evoked monochromatically, the corresponding wavelength being λ_{m}. If the visibility function $V(\lambda_{\mathrm{m}}, \lambda)$ for that level of subjective brightness is known, it is possible to find the luminance $L_E = K \int E(\lambda)\, V_{\mathrm{p}}(\lambda)\, \mathrm{d}\lambda$ of a radiation $E(\lambda)$ for which $\int E(\lambda)\, V(\lambda_{\mathrm{m}},\lambda)\, \mathrm{d}\lambda = E_{\mathrm{m}}$, giving the same brightness impression as E_{m}. This is achieved by a graphical method which can be incorporated in a slide rule for low-level photometry. Luminance scales for different relative energy distributions are easily obtained and, after chosing one of them as a standard of subjective brightness, other luminances may be corrected according to the standard scale chosen. Luminances for monochromatic radiations of wavelength λ may be obtained, in the same way. When plotting these versus λ for a number of values of E_{m}, Bouma's diagram is obtained, in which the data for composite radiations may also be plotted. An analysis is given of Weaver's data regarding $V(\lambda_{\mathrm{m}}, \lambda)$, and the various functions discussed in the preceding sections are calculated using these data. Finally the function $E_m(\lambda_m)$, as obtained from these calculations, is checked with other data from literature.

1988: J. J. Went and H. P. Wijn: The magnetisation process in ferrites (Phys. Rev. **82**, 269-270, 1951, No. 2).

Below the gyromagnetic resonance frequency the initial permeability of ferrites is, in general, independent of frequency. From this it has already been concluded that the magnetisation at very low field strengths must be governed entirely by rotational processes. Measurements at higher field strengths, (up to $\mu_0 H = 10^{-4}$ Wb/m^2 ($H = 1$ Oe), and as a function of frequency up to 1800 kc/s, have shown that a second mechanism exists, which (*a*) is frequency-dependent at frequencies below the gyromagnetic resonance frequency, (*b*) takes place at a lower field strength as the density of the ferrite is greater, (*c*) is irreversible. The additional process is ascribed to irreversible displacements of the Blochwalls.

1989: F. A. Kröger and P. Zalm: The origin of multiple bands in luminophors activated by divalent manganese (J. Electrochem. Soc. **98**, 177-182, 1951, No. 5).

The two principal theories which have been advanced for the explanation of the appearance of multiple bands with manganese-activated luminophors, particulary the silicates and germanates, are the "cluster theory" and the "co-ordination theory". The former accounts quite naturally for the dependence of the red/green ratio on the Mn concentration, whereas the co-ordination theory requires both the effect of co-ordination and concentration quenching (or possible some other factor) for a self-consistent picture. On the other hand, the effect of Be in promoting the red emission is quite naturally explained by the co-ordination theory, whereas the cluster theory requires an additional postulate concerning the effect of Be in promoting clusters. Taking these and other considerations into account, the cluster theory seems to present the simpler and more satisfactory picture. In the cluster theory, the variation of luminescence observed is attributed to the interaction between manganese ions belonging to clusters and it seems that this influence must affect mainly the lifetime of the various excited states. It seems probable that the fluorescence bands are due to electronic transitions from various excited states to one single ground state and that several different bands may originate in one centre. The variation of fluorescence with temperature can be explained on the basis of this picture.

1990*: P. Cornelius: Kurze Zusammenfassung der Elektrizitätslehre. Eine Einführung des rationalisierten Giorgischen Maszsystems (Vienna, Springer Verlag, 1951; VI + 89 pp., 11 figs, 7 tables). (A short survey of the theory of electricity, an introduction of the rationalized Giorgi system of electromagnetic units; in German).

For the contents of his booklet the reader is referred to abstract No. **1803***. The German edition is extended with a chapter on the difference between field strength and induction in the electric and the magnetic field and further contains an "addendum", in which the author deals briefly with possible objections of physicists regarding his way of exposition. An English edition is in preparation.

1991: J. L. Snoek: Conférence sur les propriétés magnétiques des ferrites (J. de Physique Radium **12**, 288-238, 1951, No. 3). (Conference on magnetic properties of ferrites; in French).

Survey of magnetic properties of ferrites, in which the following points are discussed: Fe_3O_4, ferrites XY_2O_4, ferrite technology, normal and inverted spinels, Néel's theory of antiferromagnetism, variation of saturation magnetisation with temperature, magnetic resonance, magnetic crystal energy, magnetostriction, dispersion and absorption at high frequencies. To this paper a report of E. W. Gorter has been added, which deals with an experimental illustration of the influence of the angle Fe-O-Fe on the indirect exchange energy.

1992: A. Bierman and W. Hondius Boldingh: The relation between tension and exposure times in radiography (Acta Radiologica **35**, 22-26, 1951, No. 1).

Experiments, carried out under widely varying conditions, have proved that the density of a radiograph remains practically constant when varying kV and mAs in such a way that the value of the product (kV)^5 × mAs is kept constant. Application of this rule does not give intolerable deviations in the whole field of diagnostics.